

Specific targeted research project  
IST call 4  
FP6-2004-IST-4

**EUropean iNfrastructure fOr legal information Modelling,  
Interoperability and Access**

**EUNOMIA**

**Date of preparation:** March 21<sup>st</sup>, 2005

Participant no.	Participant name	Participant org. short name
1 (coordinator)	Consiglio Nazionale delle Ricerche	CNR
2	Leibniz Center for Law, Faculty of Law, University of Amsterdam	LCL
3	Centro nazionale per l'Informatica nella Pubblica Amministrazione	CNIPA
4	Institute for Language and Speech Processing	ILSP
5	IUS Software	IUS
6	Research Center for ICT Law, Faculty of Law, University of Pécs	IKJK
7	Center for Sprogteknologi, University of Copenhagen	CST
8	Research Institute for Linguistics of the Hungarian Academy of Sciences	RIL-HAS
9	IRION Technologies	IRION
10	Athens Bar Association	ABA
11	Greek National Printing House	GNPH
12	Office for the Electronic Publication of Legal Data, Federal Office of Justice	COPIUR, FOJ
13	SWORD Technologies S.A.	SWORD

**Name of the coordinating person:** Pierluigi Spinosa  
**e-mail:** [spinosa@ittig.cnr.it](mailto:spinosa@ittig.cnr.it), [francesconi@ittig.cnr.it](mailto:francesconi@ittig.cnr.it)  
**fax:** +39-055-4221637

**Table of contents**

<b>STRATEGIC OBJECTIVES ADDRESSED .....</b>	<b>3</b>
<b>PROPOSAL ABSTRACT .....</b>	<b>3</b>
<b>B.1 SCIENTIFIC AND TECHNOLOGICAL OBJECTIVES OF THE PROJECT AND STATE OF THE ART..3</b>	
MOTIVATION AND OVERALL OBJECTIVE.....	3
SCIENTIFIC AND TECHNOLOGICAL OBJECTIVES .....	4
1) <i>Development of standardized, pan-European models for the representation of normative documents</i> .....	5
2) <i>Tools promoting the adoption of the European standards</i> .....	7
3) <i>Development of services upon legal standards</i> .....	9
STATE OF THE ART .....	10
<b>B.2 RELEVANCE TO THE OBJECTIVES OF THE IST PRIORITY .....</b>	<b>12</b>
<b>B.3 POTENTIAL IMPACT .....</b>	<b>15</b>
STRATEGIC IMPACT .....	15
REINFORCING COMPETITIVENESS .....	15
SOLVING SOCIETAL PROBLEMS .....	15
INNOVATION-RELATED ACTIVITIES .....	16
EXPLOITATION PLAN .....	16
DISSEMINATION PLAN .....	17
ADDED- VALUE.....	17
OTHER NATIONAL AND INTERNATIONAL RESEARCH ACTIVITIES .....	18
B.3.1 CONTRIBUTIONS TO STANDARDS .....	19
<b>B.4 THE CONSORTIUM AND PROJECT RESOURCES .....</b>	<b>20</b>
CONSORTIUM DESCRIPTION .....	20
B.4.1 SUB-CONTRACTING .....	38
<b>SCHULTZ.....</b>	<b>38</b>
<b>DEPARTMENT OF STATISTICS, OF THE UNIVERSITY OF FLORENCE.....</b>	<b>38</b>
B.4.2 OTHER COUNTRIES .....	41
<b>B.5 PROJECT MANAGEMENT.....</b>	<b>42</b>
ORGANISATION AND MANAGEMENT STRUCTURE.....	42
STEERING COMMITTEE .....	42
TECHNICAL COMMITTEE .....	43
WP LEADERS .....	43
CONFLICT RESOLUTION AND RISK MANAGEMENT .....	44
QUALITY CONTROL AND RISK ASSESSMENT .....	44
INTRODUCTION .....	46
WP1: MANAGEMENT AND COORDINATION.....	46
WP2: STANDARDIZED REPRESENTATION OF LEGAL DOCUMENTS.....	46
WP3: DESIGN OF AUTOMATISMS FOR LEGAL DOCUMENT ANNOTATION .....	48
WP5: INTEGRATED WORKBENCH FOR THE EDITING OF NORMATIVE TEXTS .....	51
WP6: METHODS AND TOOLS FOR THE QUALITY CONTROL OF NORMATIVE TEXTS.....	53
WP7: MULTILINGUAL SEMANTIC SEARCH AND RETRIEVAL OF NORMATIVE DOCUMENTS .....	53
WP8: AUTOMATIC RECONSTRUCTION OF THE NORMATIVE TEXT IN FORCE .....	54
WP9: EVALUATION.....	54
<b>WORKPACKAGE LIST.....</b>	<b>61</b>
WORKPACKAGE DESCRIPTION .....	65
<b>B.7 OTHER ISSUES .....</b>	<b>82</b>

## **Proposal summary page**

### **EUropean iNfrastructure fOr legal information Modelling, Interoperability and Access EUNOMIA**

#### **Strategic objectives addressed**

2.4.9: ICT research for Innovative Government

#### **Proposal abstract**

In today's information society, the spread of legal information at the national and trans-national levels calls for an improvement of the overall quality and accessibility of normative documents generated within different legal systems across the EU. The idea driving the EUNOMIA project is that normative data quality can be improved, and its management facilitated, by **enriching normative documents with standardised meta-data describing their formal structure and semantic content**.

To this end, EUNOMIA intends to address three main scientific and technological objectives:

- a) the definition of a *shared, pan-European description model* for modeling the formal structure and semantic content of normative documents, combined with *multilingual ontologies* for representing the knowledge expressed therein and preserving at the same time multilingualism and diversity. A common representation model for normative data will foster interoperability among different legal information systems in different countries, bridge differences between different national legal systems;
- b) the *development of mechanisms for applying the model automatically*, most of them relying on natural language processing and semantic web techniques, in order to facilitate the spread of the proposed model by reducing the need for manual, specialized work. In particular EUNOMIA will develop a law drafting environment for normative documents working on the standards proposed by the project. Similarly it will improve existing technology applied to semantic cataloguing and indexing of structured and unstructured normative information that will be provided as a dynamic source ready for sharing;
- c) on the basis of the description model and the automatic tools developed, the *implementation of advanced services for normative data management*, such as reconstruction of the normative text in force, text quality control, multilingual and cross-lingual indexing and retrieval. These services intend to demonstrate how the common representation model and associated tools proposed by EUNOMIA can have an impact on facilitating administrative processes, and can enable better communication among legal practitioners and towards European citizens, ensuring easy access to normative information both by specialised and non-specialised users.

### **B.1 Scientific and technological objectives of the project and state of the art**

#### **Motivation and overall objective**

In today's information society, the spread of legal information at the national and trans-national levels calls for an improvement of the overall quality of normative documents generated within different legal systems across the EU.

Legal information data need to be suitable at least for the following requirements:

- To guarantee interoperability among different legal information systems in different countries;
- To enable a straightforward realization of integrated services both for policy makers and for citizens;
- To enable communication among legal practitioners and towards European citizens, to make them informed and involved into the policy cycle and the decision making process;
- To provide citizens and institutions with easier access to regulative information, allowing them to cope with the heterogeneity and complexity among different legal systems in EU, and preserving at the same time multilingualism and diversity;
- To allow Public Administrations, as the main providers and distributors of legal information, to distribute and sell reliable, updated and high quality public information, and therefore decrease the cost of public services;
- To improve the information market and to create a source of business for private companies.

The EUNOMIA project aims at paving the way for the attainment of these goals through the definition of **methodologies and tools to add value to legal information** in order to improve the quality of normative documents generated within different legal systems across the EU and to facilitate their management.

The following figure (Fig.1) illustrates EUNOMIA's vision of the process of normative data enrichment and exploitation.

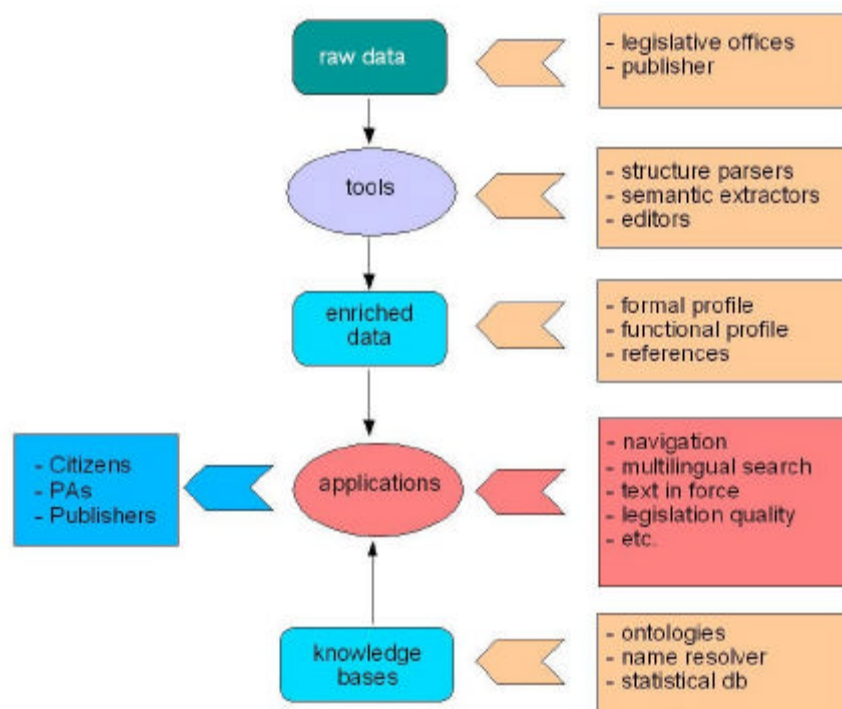


Fig. 1 Normative data enrichment and applications

### Scientific and technological objectives

Normative data need to be qualified along several dimensions: univocal identifiability inside a legal system; easy tracing of the modifications incurred during their lifecycle; possibility of ensuring easy access both to specialised and non-specialised users. In particular, the development of strategies to access legal information regardless of geographic or language barriers is a key factor for the truly sharing of legal knowledge, as well as ensuring democratic transparency and legal certainty.

The idea driving the EUNOMIA project is that normative data quality can be improved, and their management facilitated, by **enriching normative documents with standardised meta-data describing their formal structure and semantic content.**

To this end, EUNOMIA intends to address three main scientific and technological objectives, which also represent the logical pillars around which the project is structured:

- d) the definition of a *shared, pan-European description model* for modeling the formal structure and semantic content of normative documents, combined with *multilingual ontologies* for representing the knowledge expressed therein and preserving at the same time multilinguality and diversity. A common representation model for normative data will foster interoperability among different legal information systems in different countries and bridge differences between different national legal systems;
- e) the development of mechanisms *for applying* the model automatically, most of them relying on natural language processing and semantic web techniques, in order to facilitate the spread and adoption of the proposed model by reducing the need for manual, specialized work. EUNOMIA will improve existing technology applied to semantic cataloguing and indexing of structured and unstructured normative information that will be provided as a dynamic source ready for sharing.
- f) on the basis of the description model and the automatic tools developed, the *implementation of advanced services for normative data management* such as normative text-editing, reconstruction of the normative text in force, text quality control, and multilingual and cross-lingual indexing and retrieval. These services intend to demonstrate how the common representation model and associated tools proposed by EUNOMIA can have an impact in facilitating administrative processes, and enable better communication among legal practitioners and towards European citizens, ensuring easy access to normative information both by specialised and non-specialised users.

#### *1) Development of standardized, pan-European models for the representation of normative documents*

Currently in Europe different national initiatives have introduced standards for legal contents description as “NIR project” in Italy ([www.nir.it](http://www.nir.it)), “LexDania” in Denmark, “MetaLex” in Netherlands, CHLexML in Switzerland. Other similar initiatives in UK have established specific metadata models, such as the “Legal and Advice Sectors Metadata Scheme” (LAMS, [www.lcd.gov.uk/consult/meta/metafr.htm](http://www.lcd.gov.uk/consult/meta/metafr.htm)) that was developed by the Lord Chancellor's Department as part of the Community Legal Service (CLS), launched in April 2000. This is a central plank in the Government's programme of legal reform, designed to increase access to justice for ordinary people; “Justice Sector Metadata Scheme” (JSMS) development of the AGLS metadata scheme. It is designed for the use of organisations in New South Wales, and Australia, which are publishing legal materials on the Internet; UK Metadata Framework (UKMF) whose objective is to describe all resources within the government sector so that policy-makers have access to the resources on a particular policy issue, regardless of the department to which those resources belong.

These experiences are mature enough for coordinating the various national initiatives and porting the national standards at a European level, in order to avoid fragmentation and proliferation of diverse national standards. Besides the other advantages, a shared standard representation model for normative documents will ensure better interoperability among the different legal systems as well as promoting knowledge and awareness of citizens in an enlarged normative landscape.

The definition of a pan-European representation model for normative documents is a scientific objective *per se*: while Legal Theory provides multifaceted, abstract models of normative knowledge, in a domain such that of law, where the dependency on language is so strong, a descriptive model of normative documents is still missing. One of the scientific goal of EUNOMIA

is to make it explicit the intersection between textual and conceptual knowledge and to define and organize the integration between linguistic entities (KE) and formal, ontological models.

An overall structure of the European-wide legal domain textually-based will provide a better arrangement of *concepts*, thus better fitting those cases in which simple lexical equivalence does not account for the complexity of relationships between different legal systems.

The definition of a general standard framework, capable to cope with both syntactical and semantic aspects will provide an effective added value not only for citizens and P.A., but also for the market: knowledge of legal aspects is a prerequisite for entering into business relationships with companies settled in a different country. By enlarging access capabilities to the European legal system, we expect to create enabling conditions both for cross European partnerships and to increase interests for abroad investments into European activities.

The development of such a standard depends, as a first step, on the investigation of state of the art on existing formal and semantic representation formalisms adopted in the different EU countries. A comparative study among different models will be carried out in order to come to a shared model at the European level.

EUNOMIA foresees three levels of modelling normative documents:

- At the higher level, any document needs to receive a description able to univocally identify it. This description must also be able to provide a way to express references in a reliable and long-term way, independently of the physical locations of the actual documents.
- Normative documents can be modelled according to their structural components, as references to other laws, formatted text-embedded relevant entities (tables, lists, etc.), and, most importantly, those elements specific to normative texts, such as heading, preamble, sections, articles, paragraphs, etc. This description constitutes the *formal profile* of a normative text.
- Finally, normative documents can be modelled according to their content or semantics: what is the subject matter addressed, what are the *normative provisions* expressed. This level of formal representation constitutes the *functional profile* of a normative text and can be further specified at two levels of detail:
  - General: representing general information about the normative act contained in the document, such as subject classification, publication date, issuing body or organisation, relationships among acts.
  - Analytic: representing finer-grained information about the semantics of the normative documents, where by semantics we mean the particular provisions expressed (i.e. an obligation, a permission, a penalty, etc.).

These levels of representation naturally translate into metadata. Metadata in law texts can be considered as a means for providing legal data with a higher level of quality according to a semantic point of view. In particular, the insertion of analytical metadata permits different applications. For instance, they allow citizens to access law documents from a semantic point of view, thus facilitating the retrieval of specific categories of norms related to specific categories of citizens. Analytical metadata can also be used in the drafting phase giving the drafter the possibility to insert metadata annotation and also of drawing up new law documents starting from their logical structure, viewed through the provision model, using an editor able to use metadata in the phase of building up a text. Moreover their presence in law texts allows to perform analyses concerning the coherency of the legal system.

Finally, the insertion of the Amendments metadata will allow the possibility to carry out automatically a consolidation process able obtain the in-force text and rules (see WP8).

The presence in the Consortium of partners such as Public Administrations, normative data producers and professional organizations will serve to ensure that the proposed representation formalisms actually comply with user needs; at the same time, it will help checking their usefulness and spreading their use among final users.

### 1.1) Ontologies

Legal discourse in its various forms (legislation, case-law, legal literature) depends on language, and therefore legal knowledge is closely related to language use within the legal domain. Legal discourse can never escape its own textuality, which could lead to the postulation that there is, as in other terminological domains, a relatively high level of dependence between legal concepts and their linguistic realization in the various forms of legal language. The strong connection between abstract concepts and their linguistic realization implies a double perspective in structuring legal information. On the one hand, there is ontological structuring in the form of a conceptual model of the legal domain. On the other, there is a vocabulary of lexical items that lexicalize concepts and are associated with specific linguistic information.

Semantic lexicons are considered *lightweight ontologies*, linguistic expansions of the description of a way of perceiving reality. It is possible that a lexicon with a semantic hierarchy might serve as the basis for a useful ontology, and an ontology may serve as a grounding for a lexicon. This may be so in technical domains, in which vocabulary and ontology are more closely tied than in more general domains.” (Graeme Hirst, *Ontology and the Lexicon*, in (Staab and Studer eds.) *Handbook on Ontologies*, Springer, 2004, p.14). These possible migrations and interdependencies between ontological and linguistic knowledge indicate that there is a trade-off between linguistic ontology design and abstract ontology design. On the formal end of the continuum are maximally language independent ontologies reflecting the structure of the domain in question. On the linguistic end there is a complete ontological dependency on domain-specific lexis and grammar. language.

Thus, both lexical and formal ontologies represent the privileged means for modelling normative and legal knowledge. The project will privilege the re-use and integration of existing proposals in the field of multilingual lexicons for law and of legal ontologies, for instance multilingual localisation of legal WordNets as developed in the Lois project (EDC 22161).

The WordNet-like lexicons can be considered as standard in the paradigm of computational lexicons. Legal WordNet acts as a support in legal information searching, as leaf concepts are structured according to taxonomy and semantic relationships based on linguistic rules. In order to remove terminological and conceptual ambiguities and to face multilingual information handling, the high level concepts in the lexicon need to be framed and organised via a *Core Legal Ontology* (CLO).

There are a number of formal legal ontologies available that approach formal modelling from different perspectives, such as the Functional Ontology (Valente, 1995) and the Frame Based Ontology (van Kralingen, 1995). Furthermore, the Dolce foundational ontology ([www.loa-cnr.it/DOLCE.html](http://www.loa-cnr.it/DOLCE.html)) provides a logically coherent system of top-level concepts such as substance, endurant and process. The Legal Core Ontology organises juridical concepts and relations on the basis of formal properties defined in DOLCE at a more specific level.

### 2) Tools promoting the adoption of the European standards

EUNOMIA is aware that mechanisms and algorithms for automatically and semi-automatically analysing normative documents are needed in order to make the adoption of the suggested European standards easier.

The project thus intends to develop a number of tools dealing with both the formal and the functional profile of normative documents.

Natural language processing, syntactical parsing techniques and machine learning procedures will be used to transform legacy legal contents into the formal representation and the semantic description of the documents, while ontologies and semantic networks will be associated to metadata in order to help legal document access and to assist document navigation and processing.

The following tools are envisaged:

- a) *Cross-reference and Structure Parsers*: Mechanisms for legacy content handling with respect to the structural standard models suggested by the project will be developed. In particular, two types of parsers will be implemented:
- A parser for the detection of cross-references among normative documents and for constructing the related URNs;
  - Parsers for detecting the formal profile of a normative text, producing its XML description related to the European legal formal standard.
- b) *Natural Language Processing for semantic analysis of normative documents*:  
The automatic analysis of the semantic content of normative texts requires tools able to both assigning a portion of a normative text to a provision type and rendering the conceptual roles relevant for a particular type of provision explicit.  
An innovative aspect of EUNOMIA is its accent on natural language techniques to support the building of formal models of normative documents. The availability of robust human language technologies is an essential pre-requisite for the development of algorithms and tools capable of automatically detecting the functional profile of normative texts. On the other hand, the design and implementation of algorithms for fine-grained semantic analysis of documents is at the forefront of research in computational linguistics.  
The project intends to avail itself of current Natural Language Processing techniques for the identification, extraction and modelling of the normative content of legal documents. The Consortium encompasses some of the most prominent research institutes in Human Language Technology and will exploit this expertise for the development of advanced tools for the automatic and semi-automatic semantic analysis of documents for the Italian, Greek, Dutch, Hungarian, and Danish languages.
- c) *Semantic web methods for normative knowledge representation and management*:  
Another scientific objective of the project is the development of methods and tools supporting the multilingual conceptual indexing of normative documents, which is by itself a prerequisite for the development of advanced search and retrieval systems. This objective will be achieved by extending current state of the art in ontology learning and re-use, harmonisation and integration of core and domain legal ontologies. EUNOMIA will explore the automatic building-up of a specific domain ontology. Usually the construction of ontologies is focused on specific domains so that they can be, in acceptable time, effective and detailed: for the project a *consumer protection* domain is proposed to be investigated. The integration of specific legal ontologies within different services developed on legal standards is a key issue of the project. The combination *provisions-ontologies* allows to join two different and complementary knowledges: the “normative knowledge” (thanks to the *provision-argument model*), and the “general knowledge” contained within the ontologies. The use of the combination *provisions-ontologies* will allow to link legal information services to well-grounded legal concepts: for example ontologies will be used both to guide users in legal information access using normalized terms and concepts, providing also a support to multilingual legal information access, as well as in the process of production of new bills, allowing their design from the conceptual point of view.
- d) *Text editor for law-makers*: An editor operating within the URN and XML-DTDs or XML-Schema framework established by the project will be developed. The editor will be able to work in two situations: it will be designed to process legacy law texts, as well as to assist the drafting of new texts. In both these working situations the editor will handle the formal as well as the functional profile of a law text, using both manual and automatic facilities. The semantics expressed through analytical metadata will play a central role also in the drafting process of new acts, as a help while expressing provision texts and as a guide in organising the text itself, to increase the general technical level of legislation as expected in the objectives of the project.



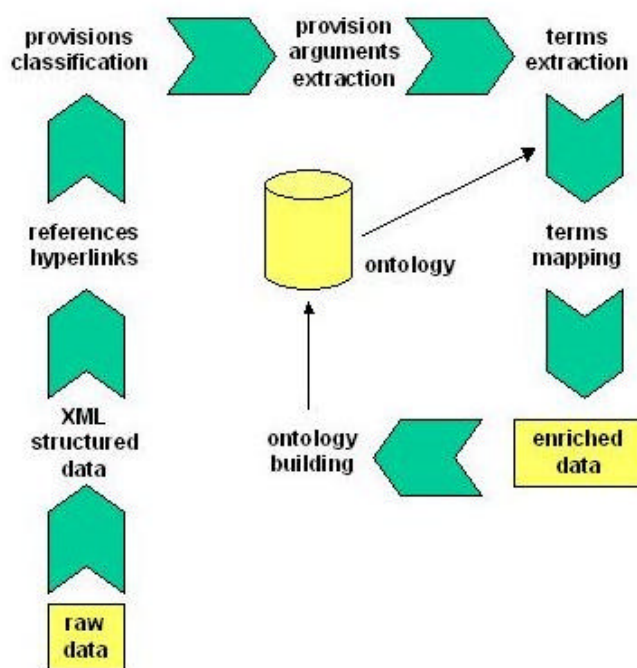


Fig. 2 Processing flowchart

### 3) Development of services upon legal standards

On the basis of the suggested standards high quality services both for law makers and for citizens will be implemented.

The adoption of European standards will permit interoperability among legal information systems of European countries with regard to the possibility of cross-referencing law documents in a stable way, independently of document physical locations.

Similarly, they will permit the development of integrated search and retrieval services on different country legal information systems. In this context multilingualism is a crucial problem in law, as cross-lingual correspondences between legal terms cannot be established on the sole basis of meaning equivalence. Thus a multilingual source of metadata for semantic annotation needs to be driven by reference to a common core of legal notions, understood in unambiguous way.

The use of metadata suggested by the project and of concepts derived from legal ontologies will support cross-language retrieval facilities preserving the identity of each legal system; moreover, it will provide effective semantic search facilities on norms of different countries, allowing the possibility to build-up comparative legal services among different legal systems.

EUNOMIA will address the technological objectives of implementing the following three services making use of normative data enriched with the proposed standardised representation model:

3.1) *Multilingual and cross-lingual retrieval* (semantic indexing and retrieval based on conceptual roles (WP3) and going beyond exact keyword matching (WP4) based on ontologies etc. The latter allowing multilingual and cross-lingual searches (WP4)).

3.2) *Normative text consolidation* (automatic reconstruction of the text in force in consequence of amendments);

3.3) *Models for the assessment of Legal Text Quality* will be implemented by the definition of evaluation-parameter based on proper content information extracted from legislative texts metadata. The goal is to provide ways of assessing adherence to technical standards, impact on the whole legislative system and possible socio-economic consequences.

**State of the art**

Currently in Europe different national initiatives have introduced standards for legal contents description as “NIR project” in Italy ([www.nir.it](http://www.nir.it)), “LexDania” in Denmark, “MetaLex” in Netherlands, CHLexML in Switzerland. Other similar initiatives in UK have established specific metadata models, such as the “Legal and Advice Sectors Metadata Scheme” (LAMS, [www.lcd.gov.uk/consult/meta/metafr.htm](http://www.lcd.gov.uk/consult/meta/metafr.htm)) that was developed by the Lord Chancellor's Department as part of the Community Legal Service (CLS), launched in April 2000. This is a central plank in the Government's programme of legal reform, designed to increase access to justice for ordinary people; “Justice Sector Metadata Scheme” (JSMS) development of the AGLS metadata scheme. It is designed for the use of organisations in New South Wales, and Australia, which are publishing legal materials on the Internet; UK Metadata Framework (UKMF) whose objective is to describe all resources within the government sector so that policy-makers have access to the resources on a particular policy issue, regardless of the department to which those resources belong. EUNOMIA project aims at harmonizing such different standards, coping also with different legal specifications of other European countries, and to produce a new legal standard as a generalization of the national ones, according to an object-oriented framework (see Fig.4 in WP2 description).

Similarly, tools promoting the adoption of national standards and to manage normative documents has been developed within national initiatives.

One of the main is IUS-TIME developed by IUS-Software. It is an XML driven Content Management System which brings extremely efficient and affordable approach to complex text based information management including processing, storage and publishing. It is able to convert structured and unstructured data into XML, providing also editing capability, data storage facilities, sophisticated full-text search methodologies and publishing functionalities.

Similar tools specifically devoted to normative drafting on national standards has been developed within NIR project (NREditor developed by ITTIG-CNR) and MetaLex project (MetaLex editor developed by Leibniz Center for Law, of the University of Amsterdam).

Particularly NREditor consists of a law drafting environment supporting the Italian legal standards (URNs and DTDs) established by the NIR project.

NREditor operates in two working situations: it is designed to process legacy law texts, as well as to assist the drafting of new texts. In both these two working situations NREditor is designed to handle the formal as well as the functional profile of a law text, using both manual and automatic facilities.

For what concerns

Specifically addressed to retrieval in multilingual legal domain is *Autonomy*, is a general purpose knowledge management system. Considering its properties for its usefulness and suitability in the legal domain we note the following:

- a) It allows for monolingual language-independent retrieval, but does not cater for cross-lingual retrieval scenarios
- b) It performs indexing by choosing the ‘right’ concept according to the context (by using naïve Bayesian methods); the retrieval relies on pattern-matching at the concept level too (the user’s keyword is matched to a concept). This might work for small documents within companies in the business domain for indexing emails, memos etc. However in the legal domain, choosing automatically the ‘right’ concept is impossible; legal documents are lengthy and highly informative, i.e. they need to be partitioned to meaningful sections each one of which refers to specific concepts/topics, namely they need to be indexed at the section level rather than the document level and they need informative indexes which will go beyond one word concepts.

- c) The way in which the notion of “right” concept based indexing is implemented does not cater for word sense ambiguity problems, leading to low precision of the retrieved results while retaining retrieval recall.

Other search engines are well suited to work on metadata indexing and retrieval.

*Metafaq* by Transversal is used for answering questions posed by customers of legal firms. It has a memory engine to do so, by relying exclusively on human-generated answers that have been given to similar questions in its database (it determines similarity using concept clustering techniques).

Within CELEX (the EU legal document indexing and retrieval engine), and in all state of the art indexing engines, the indexing and retrieval of legal documents relies on metadata information assigned to each document by its creator and/or the person responsible for indexing the document in a database. This information could be one or more of the following: unique document ID, publication reference number, publication date, document type and subtype (e.g. legislation, case law, preparatory acts etc.) and manually assigned descriptive keywords (e.g. “biodiversity”, “environmental protection” etc, for a legal document on the protection of marine environment). Searching for a document in such databases relies, in its turn, on this type of information or/and user submitted keywords that may appear in the title or main body of the document. Therefore, one needs to know very well what one is looking for in order to retrieve documents quickly and effectively, or at least one needs to know (and restrict herself to) the finite set of descriptive keywords associated to the documents of the database. When searching for legal documents related to a specific topic but created and enacted in different EU countries, difficulties augment wildly and the task becomes almost impossible without help from legal experts of different nationalities.

The *TwentyOne* Search engine developed by Irion Technologies combines powerful meta search with conceptual search at a phrase level. A user query is analyzed as a concept and matched with concepts represented in a document. The textual user-queries can be narrowed down by settings XML-based filters for any published meta-data, resulting in guided-navigation search.

TwentyOne thus incorporates both the benefits of enriched meta-data search with precise conceptual retrieval in free text. Irion developed a conceptual representation and matching algorithm that allows for retrieval across languages and across expressions within languages. It handles linguistic variation in a natural and precise way. Whereas context is normally a complete document or page (Autonomy, Verity), TwentyOne takes a linguistic phrase as the context, in which it can detect specific multiword combinations and terminology. Furthermore, it uses multilingual wordnets to define and detect meanings across languages and to provide conceptual matching across languages. The generic multilingual wordnets (currently available for 9 languages).can be replaced or augmented with domain-specific wordnets. This will result in a more controlled conceptual matching.

The TwentyOne system can be used to evaluate the effectiveness of conceptual processing of legal documents. The automatically extracted meta-information can be used in the meta-search environment and the derived and tuned semantic networks can be used for the conceptual text retrieval. Cross-lingual retrieval can be supported for all languages that are linked to the English wordnet.

The TwentyOne system has a sophisticated benchmark environment that allows you to extract large amounts of queries that can be run on different versions of the indexes. By translating the queries to the query-languages, the test queries can be used to measure cross-lingual results. The indexes can then be built with and without the outcomes of the project and the differences in recall and precision can be measured with respect to the (cross-lingual) test queries. This benchmark environment can be used to measure the effectiveness of the project results, both with respect to the automatic extraction of the meta-data and with respect to the development of specialized semantic networks for the legal domain.

**B.2 Relevance to the objectives of the IST Priority**

<b>IST Objectives</b>	<b><i>Project contribution</i></b>
<p>Main objective of the IST work programme</p> <p><i>“The aim of this work is to develop information society technologies to enable European workers and enterprises, in particular SMEs, to increase their competitiveness in the global marketplace, whilst at the same time improving the quality of the individual’s working life...”</i></p>	<p>Public administrations, enterprises and citizens are burdened by a growing quantity and complexity of rules and regulative documents coming from various sources at different levels.</p> <p>This rapid development poses a challenge for the manageability of legal documents both for producers and users (professionals and citizens) of normative documents.</p> <p>Interoperability among legal information systems and cross-language legal document retrieval at the European level are the main expected results of the project.</p> <p>The provided standards and tools will be aimed at allowing, not only the retrieval of documents, but also of meaningful fragments of legal texts pertaining to the query, so as to implement effective facilities of <i>information extraction</i> based on the norm semantics. Similarly inferences and deductions will be possible, where the underlying norm model foresees them, in order to discover norms not explicitly expressed by the legislator.</p>
<p><b>Objectives of the Strategic Priority “ICT research for Innovative Government”</b></p>	<p><b><i>Project contribution</i></b></p>
<p><b>Technical objectives:</b></p>	
<p><i>“Research ... should include technologies to support the legislative and policy development process such as intelligent tools (...) to manage administrative processes and content”</i></p>	<p>The main objective of the proposal is the definition of methodologies and tools to add value to legal information in order to improve the quality of normative documents generated within different legal systems across the EU and to facilitate their management. This aim will be reached by adopting shared metadata for modeling formal and semantic structure of documents, combined with ontologies for linking their terms (even of different languages) to legal concepts of reference, preserving at the same time multilinguality and diversity</p> <p>EUNOMIA will improve existing technology applied to semantic cataloguing and indexing of structured and unstructured normative information that will be provided as a dynamic source ready for sharing.</p> <p>EUNOMIA will develop new and advanced methodologies for the automatic and semi-automatic tools for the management of new or legacy normative documents according to the suggested standards. The functionalities of such tools range from the drafting of new normative texts to the automatic reconstruction of the normative text in force, two aspects of strategic</p>

	importance in the management of administrative processes.
<b>Scientific Objectives:</b>	
<i>Research should address modelling of administrative processes using emerging ontology and semantic web languages.</i>	<p>A key issue of EUNOMIA is the integration of a two-level semantic representation, and hence indexing, of normative documents. On the one hand the documents will be modelled with reference to the regulatory content expressed therein, i.e. the particular provisions issued by the normative authorities; this representation is to be coupled with a representation of the regulated content of the documents, i.e. the entities, properties and relations addressed by the documents.</p> <p>The combination <i>provisions-ontologies</i> is a novelty and it allows to join two different and complementary knowledges: the “normative knowledge” (thanks to the <i>provision-argument model</i>), and the “general knowledge” as represented in ontologies.</p> <p>Cross-language retrieval facilities in legal context cannot only be established on the basis of meaning equivalence: therefore strategies for query conceptualizations will be implemented so that the interaction between legal languages respect the identity and peculiarities of each legal system.</p>
<i>Research should respond to public service governance requirements such as process transparency</i>	<p>Normative data need to be qualified along several dimensions: univocal identifiability inside a legal system; easy tracing of the modifications incurred during their lifecycle; possibility of ensuring easy access both to specialised and non-specialised users.</p> <p>In particular, the development of strategies to access legal information regardless of geographic or language barriers is a key factor for the truly sharing of legal knowledge, as well as ensuring democratic transparency and legal certainty.</p> <p>A particular advantage for new entered countries will be the possibility to share already tested methodologies and standard to speed up the integration process.</p>
<b>Societal and policy objectives:</b>	
<i>Public services that are provided for all</i>	The development of strategies to access legal information regardless of geographic or language barriers is a key factor for the truly sharing of legal knowledge, as well as ensuring democratic transparency and legal certainty.
<i>Promoting the widest and best possible use of ICT-based products and services by all citizens</i>	One of the key aspects of EUNOMIA is the accent posed on the use of natural language as the most natural interface through which citizens can access information. Citizen-friendly services must make use of natural language in order to access specialized information. Normative documents contain information expressed in a specialized, often jargon-like form. Citizens do not

	always possess the degree of specialized knowledge or the technical language necessary to efficiently access this information. By translating the normative meaning of a legal text and mapping juridical concepts onto both specialised and everyday terms, EUNOMIA aims at bridging the gap between everyday, unspecialised citizens' lexicons and specialised and technical lexicons.
<i>Preserving multi-linguality</i>	

## B.3 Potential impact

### Strategic impact

EUNOMIA squarely addresses the problem of finding innovative solutions to both the proliferation and fragmentation of normative knowledge at the national and supra-national levels. More in detail, the main issues to which EUNOMIA will contribute are the following:

- The adoption of the standards proposed by EUNOMIA will encourage a European harmonized process of law production, with regard to conformance to quality standards, monitoring the law procedural iter, normative impact and consolidation.
- The availability of a shared descriptive model for normative documents will help interoperability among different legal information systems in different countries.
- Normative documents enriched with standardized metadata will enable a straightforward realization of integrated services both for policy makers and for citizens.
- overall, the application of the proposed metadata to the quality check of normative documents and to the reconstruction of the normative texts in force will allow Public Administrations, as the main providers and distributors of legal information, to distribute and sell reliable, updated and high quality public information, and therefore decrease the cost of public services;
- The provided standards and tools will be aimed at allowing, not only the retrieval of documents, but also of meaningful fragments of legal texts pertaining to the query, so as to implement effective facilities of *information extraction* based on the norm semantics. In this way inferences and deductions will be possible, where the underlying norm model foresees them, in order to discover norms not explicitly expressed by the legislator.
- To improve the information market and to create a source of business for private companies (this is in line with recommendations to public institutions contained in the recent “*European Parliament and Council Directive on the re-use and commercial exploitation of public sector documents*” (Brussels 5.6.2002, COM (2002) 207))

### Reinforcing competitiveness

Advanced legal content processing will enable government agencies to significantly reduce production costs and at the same time increase end-user satisfaction by offering better quality content with more added value. Offering more for less is especially important in the context of a universal legal obligation to provide free access to public information. In the private sector, only those who can provide more for less will survive.

### Solving societal problems

- Multilingual access to data will enable communication among legal practitioners and towards European citizens, to make them informed and involved into the policy cycle and the decision making process; at the same time, it will preserve multilingualism and diversity.
- Conceptual indexing of normative documents, combined with the type of semantic representation of regulatory content as proposed by the project will provide citizens and institutions with easier access to regulative information, allowing them to cope with the heterogeneity and complexity among different legal systems in EU.
- the use or the creation of ontologies and semantic networks, using also well-established products such as EuroWordNets and their extensions to the legal domain will allow to overcoming linguistic and conceptual barriers in legal document searching and retrieving,

## **Innovation-related activities**

Innovative elements are based on a standardized extensible XML platform which allows for complete flexibility when dealing with legal content. This enables fully automated processing of unstructured legal content based on pattern recognition etc. Some of the practical benefits are semantic cross-referencing, point in time search, unsupervised classification, granular versioning.

We propose the development of software, which will go beyond current practices and will allow for fully automatic indexing and retrieval of legal documents according to their semantics, i.e., it will classify the sections of e.g. an act according to the provision type and the arguments of the norms they refer to (the legal entities involved and the regulatory context e.g. the duties of an employee). This way, the retrieval of legal documents will be semantics-based (it will rely on the actual meaning of their contents) and fine-grained (since retrieval will go down to the section level rather than merely the document level). The tool will allow for cross-lingual retrieval of legal documents, i.e. a user will be able to submit a search query in e.g. Greek and retrieve related legal documents written in different languages.

The extraction of semantic role labels will be applied to indexing/retrieval for the first time. (also cf. the ConLL workshop series in ACL conferences when referring to state of the art of the techniques that will be used for semantic role extraction etc.).

- The effectiveness and use of WordNets for specific domains in cross-lingual information retrieval. An important aspect is to see whether the conceptual indexes and the conceptual query-index matching of the TwentyOne system are also useful for a specialized domain. Currently, this has only been demonstrated for general news. We expect that the legal domain has less ambiguity, more terminology and requires a much higher degree of precision.
- TwentyOne has a very open system for building meta indexes. Any information expressed in XML can be included. An innovative aspect to this is the complexity of the meta data in this project. The meta indexes of TwentyOne are flat tables and there are no correspondences between these tables. If the project derives more object-oriented meta data, the meta data indexing and retrieval needs to be adapted to handle hierarchical dependencies between specifications.

The porting of the TwentyOne system to other European languages. The basic TwentyOne system is delivered for 6 European languages: English, Dutch, German, French, Spanish and Italian. In the MEANING project, the system was extended with Basque and Catalan. In this project, we will extend the system to: Greek, Hungarian, and Danish.

## **Exploitation plan**

EUNOMIA intends to exploit the results of the project according to the following classification:

### ***1. Resources Exploitation***

Repositories of multilingual semantic metadata: For organisations involved in exploitation, it is of the highest importance that the creators of information apply adequate mechanisms for identification and description of the information. Main issues here are the assignment of standard identifiers (ISBN, ISSN, DOI, maybe URNs) and standard metadata (Dublin Core, ONIX or ebXML). Proposals in the field of encoding schemes and content description and management (data representation, controlled vocabularies such as GCL and Eurovoc) are still few and not domain specialised.



## **2. *Software Exploitation***

Automatic classification, indexing and Searching strategies: Publishing and presentation strategies are driven for a large part by technical developments. Harvesting techniques and interoperability aspects can play an essential role.

Cross-lingual searching engine: Web search engines are an important input in the structuring of access, combining broad-brush searching strategies with user-oriented classification and metadata search facilities. They offer opportunities to bring together collections from various sources to provide value-added services.

## **3. *Services Exploitation***

Standard generation: All European member countries, and the EU itself, are committed to the development and implementation of standards. Unless providers of Public Service Information understand and adhere to these standards, the information produced will not be interoperable with other sources of information and the possibilities of re-use will be severely limited. Standards add value to and reduce the cost of collecting, aggregating and providing PSI, thereby providing best value to the taxpayers who fund it.

## **Dissemination plan**

Establishment of strategic relationships with government agencies and aggressive commercial exploitation via new and existing on-line services offered by SMEs will be the preferred methods of dissemination.

Moreover, EUNOMIA envisages the following tasks to disseminate the project results:

- Conference and seminars;
- Workshops and exhibitions;
- Public documents and articles;
- Web events;
- Marketing material.

The foreseen marketing and business dissemination actions will be directed towards the more relevant consumer and commercial categories, represented by:

- Legal Information publishers and industries, which are technical and standards developers at European level contributing to the potential exploitation of public information, as well as the potential exploiters themselves.
- On-line consumers of legal information, including both citizens and law firms
- Public administration
- Commercial companies

## **Added-value**

Notwithstanding the always increasing initiatives in the field of standards generation and ontology development, a standard metadata framework for semantic annotation for law is still missing; this could provide high added value to legal information managing and delivering both for:

### **legal web sites providers:**

- solid and clear structure for information storing
- easier and speedier maintenance of data
- re-use of existing information

–better quality of data implied by the accuracy in semantic mark-up

**legal web-sites users:**

- better accesibility of legal information and deeper precision in answering;
- user-taylorred search possibilities
- data quality assesment

**Other national and international research activities**

The project is going to take into account the current initiatives which, at national levels, have established normative standards and exploited their effectiveness (NIR project in Italy, LexDania in Denmark, MetaLex in Netherlands, CHLexML in Switzerland). Similar initiatives on legal metadata definitions will be considered: Legal and Advice Sectors Metadata Scheme” (LAMS, [www.lcd.gov.uk/consult/meta/metafr.htm](http://www.lcd.gov.uk/consult/meta/metafr.htm)) developed by the Lord Chancellor's Department as part of the Community Legal Service (CLS); “Justice Sector Metadata Scheme” (JSMS) development of the AGLS metadata scheme, designed for the use of organisations in New South Wales, and Australia, which are publishing legal materials on the Internet; UK Metadata Framework (UKMF) whose objective is to describe all resources within the government sector so that policy-makers have access to the resources on a particular policy issue, regardless of the department to which those resources belong.

### B.3.1 Contributions to standards

The lack of knowledge and control of the legal order, which depends on the scarce transparency of the legal system, represents a crucial problem for both the legislator and the citizens.

Problems of different nature arise in dealing with a non-systematic organization of legal order, from the uncertainty of the impact of new laws on the legal order in terms of coherency preservation, to the difficulties in norm accessing by both citizens and legal experts.

The use of standards able to describe normative documents helps the upkeeping of legal systems and provides facilities to access norms.

In these perspectives EUNOMIA project aims to contribute meaningfully to the definition of standards, to be shared at the European level, to describe normative documents.

As discussed in Section B1, three main standards will be proposed. They will be able to:

- 1) to identify normative documents univocally and independently of their physical locations;
- 2) to describe their physical structure (*formal profile*);
- 3) to describe their semantic structure (*functional profile*).

The definition of such standards will follow the W3C recommendations, in terms of URI/URN and XML techniques to be adopted.

The use of shared standards for normative documents at the European level will guarantee interoperability among legal information systems, as well as the possibility to develop advanced services on legal data at the European level. This will pave the way to comparative studies on European legislative production within the European integration process perspectives, as well as integrated services with the aim of enhancing the awareness degree of European citizens of the policy cycle and to make them involved in the e-governance process.

In the aims of the project, the European legal standards should also represent a reference for those countries where standards for normative documents have not yet defined, so to obtain local new standards as specializations of the European one, coping with the peculiarities of each national legal system.

EUNOMIA project also aims at developing a specific normative editor (EUNOMIAEditor) able to deal with the European legal standards that will be established by the project. EUNOMIAEditor will provide a general drafting environment that can be personalised, within an object oriented framework, by each country according to the peculiarities of each national standards (Fig.6 in WP5 description).

## B.4 The consortium and project resources

### Consortium Description

The EUNOMIA Consortium possesses the necessary know-how to develop the proposed work. The Research Institutes and Universities involved are amongst the most prominent ones both in the fields of legal informatics and legal modelling and in the field of Natural Language Processing and Semantic Web.

#### Consiglio Nazionale delle Ricerche

Consiglio Nazionale delle Ricerche (Italian Nation Research Council - CNR) is the coordinator of the EUNOMIA project. It is involved within the project with three institutes (CNR-ITTIG, CNR-ILC, CNR-LOA). Apart from other contributions to the project, CNR-ITTIG and CNR-ILC will coordinate the technical and financial aspects.

#### CNR - ITTIG - Istituto di teoria e tecnica per l'informazione giuridica

The Istituto di teoria e tecnica per l'informazione giuridica (CNR-ITTIG) is a research center of the Italian Nation Research Council (C.N.R.), a government institution. It carries out studies and research and creates applications in the field of legal informatics and computer law. Its scientific output is therefore along the interdisciplinary lines, in which the science and practice of Law meet with Information technology and with the sciences of Language and Documentation. In particular the Institute studies and applies information and communication technologies in the sector of legal documentation and legal language, legislative drafting, legal decision-making, legal education. Current research activities are carried out in relation to three main areas:

- *New technologies for Public Administrations:* Theoretical and applied study of documentary and computer methodologies and the definition of legal tools for the purpose of developing and integrating public Information Systems as well as their legal regulation. The main project at the present, is the NIR (Norme in Rete) project, aimed at developing standard methodologies for the digital publication of statutes, and at creating an Internet Portal, that allows free access to normative information by reaching the sites of the law makers themselves.
- *Information technology, science, practice and teaching of law;* study and experimentation of systems for teaching and learning law with the aid of computer tools.
- *Organisation and dissemination through telematics of legal information:* study and experimentation of systems for on line access to legal information, for the dissemination of on line legal data banks and for the management of legal documents.

#### Key personnel

**PierLuigi Spinosa** is a technological director at ITTIG and he has a background in electronic engineering. His research field regards different aspects of legal documentation, mainly legal data bases and projects of feasibility for public administrations aimed at the design of legal information systems and the publication of law on the Internet. Currently he coordinates the working group related to the specification of a identification common standard for legal documents (URN – Uniform Resource Name), and he is a member of the working group for standardisation for both cases and law within the NIR Project funded by the Ministry of Justice. He supervises the development of tools aimed at facilitating the adoption of NIR Standards (reference parser, automatic markup in xml language of legislative texts, XML Editor for legislative drafting).

**Daniela Tiscornia**, graduated from the Faculty of Law at Genova University, is Senior Researcher at ITTIG . Since 1980 she has been active in the field of 'computational law', that is the applications of IT technology to legal domain, where she has been working in the development of knowledge-based and conceptual retrieval systems. She has been involved in several Italian and

European Projects in the field of Legal Information Accessing and of Legal Knowledge Management; among others: Epsinet, *European Public Sector Information Network* (EDC-51003); Lois, *Lexical Ontologies for Legal Information Sharing* (EDC 22161), Caselex, *CaselawExchange* (EDC-11218). She is author of publications and scientific papers on the formal representation of legal norms and legal reasoning. Her present research interest is focussed on the development of ontology-based model of legal knowledge.

**Carlo Biagioli** graduated from the Faculty of Political Sciences at the University of Florence, is a senior researcher at the Istituto di Teoria e Tecniche dell'Informazione Giuridica of the Italian National Research Council. He worked in the field of documentary legal informatics, heading the ITTIG Computing Centre, and on legal knowledge engineering, studying formalisms for the representation of legal norms and co-carrying out several expert systems in the legal domain. He planned and co-carried out NIREditor, an XML Editor to support standards based legislative drafting, within the standardization process promoted by the Normeinrete national project. He is currently engaged in modelling the functional structure of legislative texts, adopting an interdisciplinary approach, leading to the perspective of ontology-driven information systems.

**Enrico Francesconi** received the Laurea in Electronic Engineering from University of Florence in 1993 and the Ph.D. in Computer Science in 1998 at the Computer Science Department (DSI), University of Florence. His research activity at DSI included, machine learning techniques for document analysis and understanding, and knowledge representation. In 2002 he joined the Institute of Legal Information Theory and Technologies (ITTIG-CNR) in Florence, where currently he is a research assistant. At ITTIG his main activities are related to legal standards, legal drafting (IT design of NIREditor, a legal drafting environment working on the standards defined within the NIR project) artificial intelligence techniques for legal document classification and knowledge extraction. He is a member of the NIR national working groups establishing the XML and URN standards for Italian normative documents.

**Pietro Mercatali** Law Degree, a researcher at the Istituto di teoria e tecniche dell'informazione giuridica (ITTIG) of the Italian National Research Council since 1976. He works in the fields of legal informatics, legal language and law-making. He is an expert in the electronic processing of legal documents and, in particular, in legimatics, a new discipline for designing and building computer tools for legislative drafting. Between 1983 and 1989, he spent periods in training in these domains at universities and research centres in the United States. He works with numerous bodies on developing awareness about and the computer-aided tools at legislative assemblies and, in general, in the public administration in Italy and abroad. Since 1997, he has taught Computer Science Applied to the Public Administration at the Political Science Faculty of Pisa and at the Naval Academy of Livorno.

**Francesco Romano** achieved his degree in law in 1996. Since 1997 he collaborates with the Institute of Legal Information Theory and Technologies of the Italian National Research Council, first as creation of the historic database for legal terms, afterwards to the second phase of "Norme in Rete" (Legislation on the Net) Project, promoted by the Italian Ministry of Justice. Currently, he is winner of a grant sponsored by Institute of Legal Information Theory and Technologies of Florence, of the Italian National Research Council, besides to work in a solicitor's office, to continue cooperating with the Institute of Legal Information Theory and Technologies of Florence of the Italian National Research Council, working to the projects "MarcaTeLi" for the Mark-up of Italian legislative texts, E-PSI NET (Europe's Public Sector Information project) and Ital-Wordnet.

**Fabrizio Turchi** is a technological senior at ITTIG and received the Laurea in Mathematic from University of Florence. His research activities are the applications of IT technology to legal domain, theoretical topics linked to logic and logic programming applied to the law design, study and use of

multimedia technologies in legal education and professional training, use of the XML language technologies for the construction of models of legal documents, design and development of Web applications.

## **CNR-ILC Istituto di Linguistica Computazionale**

The Istituto di Linguistica Computazionale del CNR (CNR-ILC), Pisa, born in 1967, was founded in 1978 as an independent CNR institute. It has been a major promoter of the notion of Language Resources (LR) as the central component of the “linguistic infrastructure”, aware also of its economical and political implications; has coordinated the major LR and standardisation initiatives in the last decade, and has often been promoter of new “paradigms” in the field. It has designed and built several types of corpora and large computational lexicons with related ontologies, has developed a complete chain of tools for very robust processing of the Italian language, methodologies and tools for acquisition of knowledge from corpora, machine learning, word-sense disambiguation, monolingual and multilingual terminology extraction, ontology acquisition and structuring, etc. The language technologies are used for several tasks: question answering, information retrieval, text mining, summarisation, etc., and within different application domains: filtering of web documents, knowledge management, preservation of cultural heritage through digital image processing and digital libraries techniques, e-government, etc.

CNR-ILC has participated to 52 EC Projects, 15 as Coordinator, and 4 National Projects as Coordinator.

It is represented in many international and national Committees, Boards, Associations, the Italian Forum for HLT at the Ministry of Communications, etc. It has many collaborations: international: 89 (29 countries); nationals: 30; industry: 23 (11 countries); with Ministries, region, public administrations, etc., it has contracts of research with industries and is active in the organisation of international conferences, workshops, summer schools, etc.

CNR-ILC is engaged in a number of activities related to eGovernment issues where language technologies play a strategic role:

- 1 the design and development of SALEM (Semantic Annotation for LEgal Management), an NLP-based system - currently used as an advanced module of the legal editor developed in the framework of the Italian national project NIR (Norme in Rete, sponsored by the Ministry of Justice for the free access by citizens to Italian jurisdiction) - to automatically tag the semantic structure of Italian law paragraphs through an integration of NLP and information extraction-inspired technologies;
- 2 the design and development of T2K (Text-to-Knowledge), a hybrid system for terminology extraction and ontology learning which: dynamically acquires terminology from domain documents, where terms are represented by single- and multi-words dynamically extracted from linguistically pre-processed texts and are selected by means of statistical measures which “weigh” their relevance with respect to the domain; performs ontology construction through the identification of “vertical” relations among terms (i.e. hyperonymy and hyponymy) and “horizontal” relations among terms inferred from distributionally-based similarities among terms. T2K was originally developed as a component in an advanced document management system for Public Administration, and is currently used in a number of national and international projects.
- 3 JURWORDNET: a semantic lexicon for the legal domain, developed as a specialized plug-in of the ItalWordNet general semantic lexicon.
- 4 The constitution and active involvement in ONTONOMOS, an Italian special interest group on Legal Ontologies.

CNR-ILC permanent staff: 34; temporary staff: about 25 junior researchers, Phd, etc.; consistent auto-financing.

## **Key Personnel**

**Nicoletta Calzolari Zamorani** Director of CNR-ILC. CNR Director of Research (equivalent to Full Professor). Working in the field of Computational Linguistics since 1972, and with a long experience in the coordination of many international, European and national projects and strategic initiatives, also in the eGovernment sector. Member and General Secretary of ICCL, Vice-President of the ELRA Board, founding member of the Italian Forum for HLT at the Ministry of Communications, Convenor of the ISO/TC 37/SC 4/WG4, chair of the ELRA Production Committee (PCom), member of the ELSNET Board, the SENSEVAL Advisory Committee, and of many International Committees and Advisory Boards, Conference Chair of LREC 2004, General Conference Chair for COLING-ACL 2006, chief co-editor of the new International Journal *Language Resources and Evaluation - LRE* (Springer). Invited speaker, member of program committee and organiser for quite numerous international scientific conferences, workshops etc., and of journals editorial boards. Various international and national evaluation activities. More than 300 publications.

**Simonetta Montemagni** is senior researcher at the Istituto di Linguistica Computazionale since 1998. Her major research interests are in Natural Language Processing parsing techniques (mainly for what concerns syntax and semantics), automatic acquisition of linguistic information from on-line dictionaries and textual corpora, corpus annotation, computational lexicography and lexical databases and ontology learning. Over the last several years, she has been involved in a number of national and international projects: among them, the PLNLP IBM project, EUROLANG, NERC, ET10/51 "Semantic Analysis ", EAGLES I and II, ACQUILEX I and II, PAROLE, MEMORIA, EUROWORDNET, SPARKLE, ELSE, UNL, POESIA, VIKEF. She received her Ph.D. in Computational Linguistics from the University of Manchester Institute of Science and Technology (UMIST), UK, in 1995.

**Claudia Soria** is researcher at the Istituto di Linguistica Computazionale since 1998. Her main current research interests and activities concern information extraction and automatic semantic annotation of texts, with applications to legal informatics. She holds a PhD in Linguistics and experience in annotation and representation of computational language resources, metadata, discourse and dialogue. Over the last several years, she has been involved in a number of national and international projects: among them, the MATE, ELSE, NITE, INTERA, and LIRICS.

### **Laboratorio di Ontologia Applicata, ISTC-CNR, Rome, Trento, Italy**

The Laboratory for Applied Ontology (LOA-ISTC-CNR) is the result of the fusion of two previously distinct ontology groups, one with the former LADSEB-CNR (Padua), and the other with the former ITBM-CNR (Rome). Following a major restructuring process within the Italian National Research Council (CNR), LOA now belongs to the Institute of Cognitive Sciences and Technology (ISTC). The group performs basic and applied research on the ontological foundations of conceptual modelling, exploring the role of ontologies in different fields, such as: knowledge representation, knowledge engineering, database design, information retrieval, natural language processing, and the semantic web. The group is characterised by a strong interdisciplinary approach that combines Computer Science, Philosophy and Linguistics, and relies on logic as a unifying paradigm. On the application side, special emphasis is given to the use of ontologies for electronic commerce, medical information systems, enterprise modelling, integration of lexical resources, and information access to the Web.

#### **Experience with EU-projects**

LOA-ISTC-CNR has a longstanding experience in taking part in several European research projects, such as IKF Intelligent Knowledge Fusion (EUREKA Project 2235), the OntoWeb-Ontology-based information exchange for knowledge management and electronic commerce ([www.ontoweb.org](http://www.ontoweb.org)) and WonderWeb – Ontology Infrastructure for the Semantic Web (<http://wonderweb.semanticweb.org>). The main objectives achieved by WonderWeb form a particularly significant know-how also for the SIPECO project, as follows:

- The development of a family of ontology languages that extend existing Web standards while maintaining maximum backwards compatibility. The resulting layered architecture will provide the necessary flexibility (standardising on a single language is unrealistic in the Web environment) while maximising interoperability. This work has already begun with the development of OIL, a Web based ontology language that extends and enriches RDFS.
- The development of a framework of techniques and methodologies that provide an engineering approach to the building and use of ontologies. In particular, techniques will be developed for the *semantic* integration, migration, reconciliation and sharing of ontologies, issues that will be particularly important in the development of the semantic web.
- The development of a set of foundational ontologies covering a wide range of application domains. Each of these ontologies will provide a carefully crafted taxonomic backbone with a sound high level structure that can be used as the basis for the development of more detailed domain ontologies. The integration of existing ontologies with foundational ontologies will also provide a principled mechanism for the semantic integration of ontologies.
- The development of the comprehensive technical infrastructure and tool support that will be required by real world applications in the Semantic Web. In particular, an ontology server architecture will be developed in order to link new and existing components in an integrated and extensible tool suite. This will include tools for editing, integrating and extracting ontologies as well as services such as persistent storage and reasoning support.

### Key personnel

**Aldo Gangemi** (1962) is Research Scientist at the Institute of Cognitive Sciences and Technologies of the Italian National Research Council (ISTC-CNR), Rome. He is part of the recently formed Laboratory for Applied Ontology. He graduated in Philosophy (Analysis of Scientific Languages) at the University of Rome in 1989. His main research interest is in the fields of ontological engineering and cognitive semantics. He is active in ontological engineering since 1992. He has been involved in several national and international projects on medical knowledge representation, ontological tools for conceptual modelling, legal and agricultural ontology construction and commercial catalogues building. He is author of more than 40 scientific papers (published in international journals and conferences) in the areas of artificial intelligence, medical informatics, linguistics, and philosophy. He gives international lectures and presentations on a regular basis and is consultant and lecturer for several companies on ontological engineering topics (domain ontology building, catalogue maintenance). He is coordinator of an international working group on legal ontologies and guest editor of a special issue of a scientific journal on the same topic.

**Jos Lehmann** (1971) is a researcher at the Laboratory for Applied Ontology (ISTC-CNR), in Rome, where he works on various themes in knowledge representation (e.g. (legal) ontology, causation, plans, collectives) as well as on various EU and national projects. In December 1995 he received a Master's Degree by the University of Milan, Faculty of Philosophy, for a thesis on the logical and epistemological aspects of the Frame Problem in Artificial Intelligence. After a couple of years in the private sector (IR department of Levi's Benelux, Sentient Machine Research), he worked as a junior researcher at the University of Amsterdam, Faculty of Law, Department of Computer Science and Law - Leibniz Center for Law (1998-2003). In March 2003 he received a Ph.D. Degree by the University of Amsterdam for a dissertation on the problem of Causation in Artificial Intelligence & Law.

**Massimiliano Ciaramita** (1968) graduated in Philosophy from the University of Rome "La Sapienza" in 1997. He then received a Master of Science in Computer Science in 2003, and a Ph.D. in Linguistics in 2004 from Brown University, Providence RI (USA). He is currently a researcher in the Laboratory for Applied Ontology (Institute for Cognitive Science and Technology, Italian National Research Council). His research interests are in statistical natural language processing and



machine learning. He works mainly on semantic classification, information extraction and learning in structured domains with ontologies and lexical resources.

### Leibniz Center for Law

*The Leibniz Center for Law* ([www.lri.jur.uva.nl](http://www.lri.jur.uva.nl)), formerly the Dept. of Computer Science & Law, of the University of Amsterdam exists since 1988 and currently employs 10 people. It is part of the Faculty of Law of the University and focuses on applying AI research and techniques for supporting legal practice en theory. LRI has great experience in the area of legal ontologies, artificial legal reasoning, standards for legal documents and actions, user oriented access to (legal) databases, and ICT in (legal) education.

It has participated in numerous national and European (applied) research projects. LRI has roots and connections with the former department of Social Science Informatics (SWI) of the same university (now "Human-Computer Studies Lab"). SWI has a longstanding experience in AIP research within a large number of national and European projects. This department currently employs about 30 people. It focuses on foundations of knowledge based system development, social aspects of information technology, and intelligent tutoring systems.

#### *Some Relevant Past Experience*

KADS Project P12 and P1098 (ESPRIT I), CommonKADS (P5248), ACKnowledge (ESPRIT II), KACTUS P8145 (ESPRIT III) on a methodology and tools for complex knowledge-based system construction. ALDUS (ESPRIT III) on automatic contract generation. CLIME P25414 (ESPRIT IV) concerned with improving access to large legal databases through the Internet, KDE (ESPRIT IV) about tools for knowledge management in legal and technical environments. *E-POWER* (IST-2000-28125): implemented a knowledge management solution by providing a method and tools that help to improve the quality of legislation by translating them to formal representation. [lri.jur.uva.nl/~epower/](http://lri.jur.uva.nl/~epower/). e-COURT (IST-2000-28199), a multimedia system for managing courtroom information, aimed at speeding up the search and retrieval of data in criminal trials by using multimedia data bases through inter- and intranet. [www.intrasoft-intl.com/e-court](http://www.intrasoft-intl.com/e-court). MetaLex, an open XML interchange standard for describing legal documents ([www.metalex.nl](http://www.metalex.nl)).

### Key Personnel

**Prof. Dr. J.A. Breuker** studied cognitive psychology at the University of Amsterdam. He obtained his PhD in 1981 at the same University on a dissertation titled: "Availability of Knowledge". Major areas of research are: experimental cognitive psychology, text understanding, intelligent teaching systems, AI, knowledge acquisition, man-machine interaction. His current work is particularly in: methodology for building knowledge-based systems, intelligent help and tutoring systems, and reasoning in regulation domains (law). Most of the research is performed in collaboration with academic institutes and industries in Europe, in particular within the framework of Esprit. Prof. Breuker holds a chair in Computer Science and Law he was awarded the ECAI-1984 price for expert systems research.

**Prof. Dr. T.M. van Engers** studied Cognitive Artificial Intelligence at the University of Utrecht. He received his PhD in 2001 for the thesis: "Knowledge Management; the Role of Mental Models in Business Systems Design ". In 1983 he started working for the Ministry of Finance, from 1990-2000 he was head of research of the department of "Artificial Intelligence and Audit Automation" of the Dutch tax and customs administration. Since 2000 he is program manager of the POWER program there. In 2003 he was appointed part time professor in Juridical Knowledge Management at the University of Amsterdam.

**Dr. R.G.F. Winkels** studied psychology at the University of Amsterdam, majoring in Artificial Intelligence. He received his PhD in 1992 for the thesis: "Explorations in Intelligent Tutoring and Help" (IOS Press, Amsterdam). In 1987 he started working as an AI researcher at the dept. of Social

Science Informatics, and in 1989 he moved to Computer Science and Law. He is now an associate professor in that department and works on artificial legal reasoning, legal ontologies, legal databases and information retrieval, and interactive learning environments.

### **Centro Nazionale per l'Informatica nella Pubblica Amministrazione (CNIPA)**

The Italian national centre for information technologies in the public administration (CNIPA) was established during 2003, merging the Italian National Authority for IT in the public administration (AIPA), which has been operating since 1993, and the Italian national technical centre. Its tasks are promoting, coordinating, planning and controlling the development of information systems within the government central organisations and agencies, through standardisation, interconnection and integration. The main objectives are better services, less cost, better communication and a wider support to the decision process within the Government. CNIPA assignments may be summarized as follows:

- Giving advise on issues related to Information Technology to the President of the Council of Minister and the Ministry for innovation and technologies during the law-making process.
- Giving technical regulations, including setting standards and criteria on planning, designing, developing, managing and maintaining the administrations' information systems and their interconnections, quality, organisation and security.
- Supporting interoperability, by supervising and coordinating projects involving several public administrations
- Auditing and evaluation of costs and benefits of the administrations' projects, and activities related to IS/IT procurement.
- Promoting education to improve IT competence in the public sector, and to control the activities developed by the administrations in this area.
- Promoting experimental projects that implement new technologies, in conjunction with several Italian and European research centres.

### **Key personnel**

**Caterina Lupo**, after the University master degree in Mathematics, worked in the United States in the field of numerical calculation applied to chemistry. Afterwards, Caterina focused her research activities on the field of textual information retrieval and of analysis and implementation of Artificial intelligence algorithms for language analysis, both in industry and Public Administration. Since 1998 Caterina has been working at CNIPA, where currently is responsible of the standardization and information representation Office. She is involved in the main Italian projects concerning regulatory and technological issues related to public data access. Caterina is the coordinator of the Italian national project "NormeinRete"; moreover, she is involved in several European initiatives related to XML representation of legal documents.

### **Institute for Language and Speech Processing**

The Institute for Language and Speech Processing ([www.ilsp.gr](http://www.ilsp.gr)) was founded in Athens, Greece, under the auspices of the General Secretariat for Research and Technology (Ministry of Development) with the aim to support the development of Language Technology. To this end, it has created a concentration of experienced scientists and the necessary technical infrastructure. ILSP has developed research activities in the fields of theoretical, applied and computational linguistics, natural language processing and engineering, computer assisted language learning, speech processing, synthesis and recognition. ILSP activities are organised in 5 scientific departments: Electronic Lexicography, Language Technology Applications, Educational Technology, Speech Technology, Machine Translation. The work undertaken by ILSP in the EUNOMIA project is to be conducted by the Department of Language Technology Applications (LTA). The Department is active in the framework of competitive national and European Union projects. European projects implemented by ILSP/LTA include: *translearn/le*, *hagis/le*, *aventinus/le*, *aquarelle /le*, *term-it/le*,

aventinus ii/le, mis/mlis, sensus/le, babeltrade/mlis, cimwos/ist, musa/ist, intera/eContent, reveal-this/ist. National projects include: dialogos /epet ii, kemes/epet ii, synopsi/epet ii, kairos/pened, Oikonomia/ pened. The Department conducts basic and applied research in the field of Natural Language Processing. Its main aim is to design computational models for natural language recognition and understanding. In particular, ILSP/LTA designs, implements and integrates human language technologies in systems and applications dealing with structured data as well as with unstructured data. Applications particularly relevant for the EUNOMIA project include language-aware Information Retrieval and Extraction, Natural Language Interfaces, Text Mining, Knowledge Management, etc. To achieve its goals, LTA builds and deploys the multi-level language resources necessary in the technology production process. The core of the approach adopted consists of developing and improving methods and techniques in the areas of artificial intelligence, machine learning, statistical and rule-based processing as well as hybrid combinations in an attempt to propose solutions that are flexible and adaptive to a range of linguistic phenomena and applications. Tools are designed and developed for: efficient structural and grammatical text annotation, shallow parsing, Named Entity Recognition, Term Extraction, coreference resolution, event recognition and their interrelations, Semantic Web oriented Information Extraction, Text Classification, and Word Sense Disambiguation.

Directly associated to the EUNOMIA project work is LTA's search and retrieval engine developed for and used by the Greek National Printing House and offered as a service to its customers. In addition, LTA is coordinating the REVEAL THIS project, where multimedia, multilingual indexing of parliamentary data released by the European Parliament and its EbS (Europe by Satellite) channel is targeted, with the objective of serving European citizens and professional experts with personalized information delivery.

### Key personnel

**Stelios Piperidis** is a senior researcher at the Institute for Language and Speech Processing (ILSP), Athens, Greece. He holds a Diploma in Electrical Engineering from the National Technical University of Athens (NTUA), and MSc and DIC degrees in Foundations of Advanced Information Technology from Imperial College, University of London. Since 1991, he has been working for ILSP, involved in European projects in the area of Language Engineering. He has been heading the Department of Language Technology Applications of ILSP since its inception in 1994. He was the Project Coordinator of the LRE/Translearn and IST/MUSA projects and the ILSP-site manager in the LE-Aventinus, Hagis, Term-IT, Sensus, MLIS-MIS, BabelTrade, IE-Aquarelle, Esprit-Naplus, EPET-Synopsis and PENED-Oikonomia projects. His research interests include statistical and deductive methods in natural language processing and understanding, machine translation and philosophy of language. He is teaching postgraduate courses at the National Technical University of Athens on "Logic and Language", "Logic Programming" and "Architectures of NLP systems". He has acted as independent expert in the evaluation and review of research projects of the European Commission. Since February 2001, he has been appointed national expert delegate, member of the Management Committee of the eContent programme of the European Union. He has published results of his work in international scientific books, journals and proceedings of many international conferences.

**Katerina Pastra** is a post-doctoral research associate at the Department of Language Technology Applications, Institute for Language and Speech Processing (ILSP). She holds a PhD in Artificial Intelligence from the University of Sheffield, U.K., an a MSc in Machine Translation from the University of Manchester, Institute of Science and Technology (UMIST), U.K and a BA in Greek Literature and Linguistics from the University of Athens, Greece. She has worked on intelligent, semantics-based indexing and retrieval for the FP5-CONCERTO project, for the UK-funded project SOCIS, which also involved ontology building for governmental bodies such the police, and she now works on the FP6 project REVEAL THIS which also involves intelligent multimodal indexing and retrieval. She has lectured on Usability Evaluation methods and metrics and she has organised

an international workshop on the development and reuse of evaluation metrics for language technology. She is the author of a number of publications on intelligent indexing and retrieval techniques, one of which has won a distinction by the British Computer Society. She is a member of the American Association for Artificial Intelligence and of the Association for Computational Linguistics.

**Vassilis Manios** is a legal information systems expert who will be working for EUNOMIA in ILSP. He has a BSc in Mathematics from the University of Athens and he is a PhD candidate in Informatics and Telecommunications at the same university. He has been working for many years as the administrator of the Athens Bar Association's Information System and Legal Data Bank in the development of which he has actively participated. He has been involved in a number of EU-funded projects and initiatives related to legal expert systems and education such as TELEMATICS, EUROFORM, and EQUAL. He has been a project manager in a number of nationally-funded projects for providing clerical support for legal professionals. He has also lectured on legal expert systems and the automatic processing of legal documents at the Informatics Department of the University of Athens, and to the Law School of the same university and at other institutions all over Greece. He has conducted research on the analysis and structure of the Greek national legislation and case-law and has developed a legal thesaurus. He has published a number of articles in legal scientific reviews and daily newspapers.

## **IUS Software**

IUS SOFTWARE d.o.o., Ljubljana, Slovenia, is a growing organization of over 35 employees (lawyers, IT professionals, salespeople and others). What we do: We specialize in development and integration of legal and business information systems. This includes content, knowledge and document management, processing and publishing over the Internet. What we provide: Comprehensive Slovenian and EU legislation (via Celex), up to date case law, customer-oriented research service as well as many other information sources relevant to both legal and business communities.

Major products and projects:

IUS-INFO is the only on-line legal and business information system in the Republic of Slovenia which contains up to date legislation and case law in its original form, catalogued and fully searchable. In addition, IUS-INFO provides access to the Official Gazette indexes. IUS-INFO is quickly becoming the de-facto standard in the Slovenian legal information market and has the potential to achieve the same status in the region and beyond. This is illustrated by the fact that all the major state institutions, government agencies, highest courts and multinational companies already rely on IUS-INFO on a daily basis.

IUS-INFO for the Supreme Court and other courts in Slovenia is a mirror site (publishing server) of the commercial IUS-INFO on-line legal and business information system. It runs within the government operated network which interconnects all the institutions including most Slovenian courts. EURO IUS-INFO is an on-line legal and business information system which contains up to date EU law. We process Celex data received directly from Brussels and publish it once a week as it becomes available. The content is fully searchable and cross-referenced.

IUS-TIME is XML driven Content Management System which brings extremely efficient and affordable approach to complex legal information management including processing, storage and publishing.

IT development and system integration represents a continuous process at IUS SOFTWARE. Making our internal business processes more efficient as well as helping numerous clients achieve excellence in content management is our highest priority.

## **Key Personnel**

**Igor Korelic** is the head of IT and is the person behind IUS-TIME -advanced legal content management system. Mr. Korelic has a deep understanding of what it takes to make legal content

usable in the e-world. Prior to joining IUS Software, he worked for some of the biggest Slovenian corporations. He holds a degree in computer science.

**Alex Petrovic** is a manager, systems and projects department. He was involved in designing and implementing the first EU law portal in Slovenia - EURO IUS-INFO. Mr. Petrovic specializes in all aspects of legal IT and has international experience. He has obtained his law degree in England and completed postgraduate studies in applied IT in Canada.

### **Research Center for ICT Law, Faculty of Law, University of Pécs**

The Research Center for Information and Communications Technology Law (IKJK) has been established in March 2004 at the Faculty of Law of University of Pécs. The IKJK deals with research, education and advising in the field of social and legal issues of information and communication technology. Supporting of legislation and legal practitioners with results of recent research is also considered as one of our significant tasks.

The researches of the Research Center are concerning information technology law, telecommunication law and media law. The main research themes are: Data protection and freedom of information, Electronic governance and electronic (public) administration, Copyright of software and other digital contents, Electronic commerce, Electronic payment and other transactions, Electronic signature, electronic documentation, Information security, Computer related crime, Regulation of telecommunication, Media Law, Legal expert systems, Electronic legal databases.

### **Key Personnel**

#### **Zsolt György Balogh**

Dr. Balogh received his degrees as a lawyer and applied mathematician in the Faculty of Law of University of Pécs and in Faculty of Science of University of Szeged. He is working as a university lecturer since 1988, recently heading the *Department for Information and Communications Technology Law* of Faculty of Law, University of Pécs (UP). He gained experiences in information rights from 1995 to 2001 as advisor of the parliamentary commissioner for data protection and freedom of information. Dr. Balogh controls the PhD program for Information and Communication Technology Law of the Faculty of Law of UP. He has been established the *Research Center for Information and Communications Technology Law*.

Dr. Balogh is a member of several international organizations, as the British and Irish Legal Education Technology Association (BILETA), International Association for Artificial Intelligence and Law (IAAIL) and of the advisory board of the International Review of Law, Computers and Technology.

Research field: legal environment of information society; law of telecommunications; information policy of state; protection of personal data; freedom of information; legal protection of computer software; electronic commerce and digital signature; computer related crime; legal databases and legal expert systems;

#### **Balázs Rátai**

Dr. Rátai graduated at the ELTE University Faculty of Law in 2000. From 2000 to 2004 he was working for the Communications Authority of Hungary. During that period he was involved in various regulatory activities focusing on the creation of the legal and regulatory basis of information society in Hungary.

In 2002 he became the special advisor of the president of the Communications Authority of Hungary responsible for IT related supervisory activities (IT security and quality certification, authentic electronic archivation, registered e-mail service, use of cryptography in private sector). As a special advisor he also acted as a contact person of the Communications Authority keeping correspondence with various Hungarian and international professional organizations.

Since his graduation dr. Rátaí is taking part in the Ph.D. program of the ELTE University, Faculty of Law. The main focus of his research are the theoretical foundations for the formal modeling of legal texts and the examination of the relationship between legal regulation and other forms of regulation (e.g. standards, specifications, different self-regulatory approaches). He is also teaching legal dogmatics, philosophy of law and political philosophy at ELTE University, Faculty of Law, Department of Legal Philosophy.

### **Péter Dévényi**

He is a Ph.D. student at the Ph.D. School of the University of Pécs. His main professional activities are at the Hungarian Ministry of Justice, and a PHARE Traineeship at the DG Agri.

### **Irion Technologies (Irion)**

Irion Technologies is a software company that develops language-based information retrieval systems for Internet companies and for software houses. Irion combines advanced language technology with the best statistic and heuristic approaches to information retrieval. Current Irion Products are: language-identification, multilingual semantic network, summarization, classification, morphological analysers, shallow parsers, named-entity recognition, cross-lingual search engines and dialogue systems. Currently, all language technology is available for and across 6 European languages.

Irion started in 2000 and spans a team of 8 developers and computational linguists. Irion is supported by a group of experts in IR and language technology, from TNO, Van Dale and a number of universities. Funded by TNO, Van Dale and three established venture capital companies, Irion thus provides the sound basis of established international companies. Irion is interested in language technology that brings information retrieval closer to understanding and meaning. Classification and multilingual semantic networks play an important role in this. Irion recently participated in the MEANING project (IST-2001-34460). In this project, IRION evaluated the automatic acquisition of lexical knowledge and word-sense-disambiguation in a conceptual retrieval application for the Spanish publisher EFE. With the IRION system and MEANING results, the productivity of the editorial work at EFE (i.e. finding correct pictures for news items) doubled. Within the project IRION is interested in measuring the effectivity of customized domain wordnets for a specialized domain and the possibilities of combining rich meta-search with conceptual search.

### **Key personnel**

#### **Piek Vossen**

Chief Technology Officer at Irion Technologies. He has a B.A. and a Ph.D. at the University of Amsterdam. He has published many articles in the field of computational lexicology and Natural Language Processing. He worked as a senior researcher at the University of Amsterdam in several EC projects (Acquilex I and II, Sift, EAGLES, and EuroWordNet I and II) and national projects (Links, Like). He was the site manager for Amsterdam in the Sift project and the coordinator of EuroWordNet, in which the University of Amsterdam was the main contractor. He recently participated in the MEANING project for IRION. He organised several workshops on wordnets (EACL/ACL in Madrid 1997, EACL/ACL-Senseval in Toulouse 2001) and 3 conferences (Euralex 1998 conference in Amsterdam, the 1st WordNet conference 2002 in Mysore, India and the 2<sup>nd</sup> WordNet conference 2004, Brno, Czech). He has been an active member of the Ansi-committee on Ontology Standards and the Eagles Lexicon Group, both involved in the standardization of ontologies, wordnets and lexical semantic resources. In 1999, he joined Sail-Labs as a senior researcher and manager, where he worked for almost 2 years on the development of customization techniques and tools for exploiting multilingual wordnets in information retrieval and classification. He is a board member of the European Association of Computational Linguistics (EACL) and a founder and president of the Global Wordnet Association (GWA).

### **Center for Sprogteknologi (CST), University of Copenhagen**

CST (Centre for Language Technology) is a research institute at the University of Copenhagen. It employs approximately 20 staff on a temporary basis, including computational linguists, lexicographers, computer scientists and engineers. CST's mandate is to carry out and promote strategic research and commercial development in language technology in Denmark. CST also acts as a consultant to individuals and organisations in introducing and making the best use of commercial language technology tools. In addition to research activities, the Centre has considerable experience in the development of a variety of HLT applications both in EU and national research projects and for commercial customers. Such applications include machine translation, development of company specific ontologies, XML mark up, development of inter-institutional term databases for the EU institutions and agencies (IATE), and controlled language and authoring tools.

The most important qualifications CST will bring to this project are: (semi)automatic ontology building, development of metadata sets, use of ontologies for information retrieval, use of XML for NLP mark up, corpus composition, collection and analysis, standards, evaluation methodologies, and NLP methods in general.

### Key Personnel

**Bente Maegaard** is currently temporarily employed at the University of Copenhagen, CST, as Director. Her main areas of expertise are computational linguistics, machine translation, evaluation methodology, dictionaries, and corpora. She was the leader of the EAGLES project on Methodologies for the Evaluation of NLP products and components, she was the coordinator of TEMAA, an EU project on evaluation methodologies for authoring tools, and has a thorough knowledge of ISO standards for software evaluation. Currently she leads a Danish national project on creating a corpus platform for multilingual parallel corpora, and a European project on Arabic language resources. She is a recognised researcher in the field and is currently the president of the European Association of Machine Translation (EAMT), and the European Language Resources Association (ELRA).

**Costanza Navarretta**, PhD, is senior researcher at the Center for Sprogteknologi, University of Copenhagen. She has twelve years experience in computational linguistics within the following topics: methodologies for building ontologies from texts, XML and XML-related technologies in NLP, discourse and dialogue structure, reference resolution, multimodal communication, computational lexicography, evaluation of NLP resources, grammar formalisation and implementation. She has participated in numerous Danish, Scandinavian and European NLP-related projects and has been teaching courses on XML and XML-related technologies for text modelling at the IT-University in Copenhagen since 2002.

**Bolette S. Pedersen** is PhD in Computational Linguistics at University of Copenhagen in 1998. Since 1998 Senior researcher at Center for Sprogteknologi, University of Copenhagen. Main research areas are computational lexicography, lexical semantics, wordnets and ontologies. Has been project manager of SIMPLE-DK concerned with the development of semantic lexica for Danish (SIMPLE) and involved with the application of semantic lexica and ontologies in content-based information retrieval (OntoQuery). Coordinator of the Nordic NORFA network SPINN on the harmonization of semantic language resources in the Nordic countries and project manager on VID on domain ontologies in industry. Project manager on DanNet – a newly initiated wordnet project in collaboration with Det Danske Sprog- og Litteraturselskab. Scientific committees: (Euralex Congress 2002 and 2004), International Congress of Language Resources and Evaluation (2002 and 2004), First Congress on Global WordNet 2002, Ontolex 2004, and 7<sup>th</sup> International Conference on Terminology and Content Development 2005.

**Anna Braasch** is senior researcher at Center for Sprogteknologi, University of Copenhagen. She has 18 years of research experience in computational linguistics with focus on corpus based

lexicography, lexical semantics, terminology for HLT applications and machine translation of LSP texts. Main assignments and responsibilities include projects in computational lexicography: she was project manager of LE-PAROLE-DK lexicon development (1996-98) and of the national project, STO, creating a large-scale Danish lexicon for computational linguistic applications (2001-2004). She was also involved in national standardization of lexical data collections (1997-2000), elaboration of metadata for and evaluation of NLP resources.

### **Research Institute of Linguistics, Hungarian Academy of Sciences**

The Research Institute for Linguistics of the Hungarian Academy of Sciences (RIL-HAS) is one of the leading research centres at the forefront of Human Language Technology research and development in Hungary. It has a track record of successfully completing several EU funded research projects (MULTTEXT-EAST, CONCEDE, TELRI I, TELRI II, MATCHPAD), which aimed at the application and adaptation of language technology standards to Central and Eastern European languages, as well as at the creation and development of new standards and language resources. In recent years, it has been also actively engaged in several major Language Technology projects funded by the Hungarian Government in the field of information extraction, lexical database and Treebank development.

RIL-HAS has accumulated invaluable language resources in the field of Hungarian morphology and syntax, and it has developed expertise in morphological analysis, partial parsing, named entity recognition, semantic frame analysis and verb argument structures. Since last year, the institute has been actively involved in three large scale projects with direct relevance to the present proposal. The first has just been finished and developed document classification technology for Hungarian based on the EUROVOC thesaurus, the other two are currently in their first phase and focus on the development of general ontology for Hungarian on EuroWordNet principles, as well as on its extension to the specific domains of business and telecommunication.

**Tamás Váradi.** Head Department of Corpus Linguistics Research Institute for Linguistics, Hungarian Academy of Sciences Dr. Váradi has been active in the field of language technology since the early nineties both in projects at the University of Lancaster UK, (Parsed LOB corpus, MARSEC) and in various EU funded projects at the Institute including MULTTEXT-EAST, CONCEDE, TELRI I, II, ELAN and MATCHPAD. He coordinated the preparation of the Hungarian National Corpus, a 150 million word, tagged and disambiguated corpus. He has been leading national HLT projects in the field of information extraction, tree bank development, lexical databases and finite state technology. Dr. Váradi has been commissioned as expert to review projects and project applications in the 5th and 6th framework.

**Csaba Oravecz,** MSc Researcher. Mr. Oravecz is a founding member of the Department of Corpus Linguistics, and has been active in the area of natural language processing ever since the start of his professional career. He was one of the chief developers of the Hungarian National Corpus and worked as leading researcher in practically all of the EU and home funded projects of the department (including GRAMLEX, MULTTEXT-EAST, TELRI, CONCEDE, MATCHPAD). He was co-developer of the first POS disambiguation system for Hungarian and recently coordinated the development of a large-scale Hungarian lexical database. He has also been actively involved in the document classification project as well in the two ontology projects, all funded by the Hungarian Government.

### **Athens Bar Association**

The Athens Bar Association (A.B.A.) was formed in 1894. It is a legal person governed by public law and numbers more than 18.500 members, legal practitioners of various ranks. ABA is a member of the Federation of the European Lawyer Associations (CCBE), and some of its members have served as Presidents and Vice-presidents of the federation. Its scientific and social activities are multifaceted; it produces a number of law publications and actually publishes two of the most



important law magazines in Greece, in which eminent legal scientists and law school professors contribute. Furthermore, it organizes seminars on a variety of legal topics including ones on legal information systems. It has participated in a number of EU-funded projects and initiatives related to technological support for law professionals and the education of the latter on such issues (EUROFORM, INNOVATIVE PROGRAMS, TELEMATICS, EQUAL etc.). One of the most important social activities of ABA is the formation of “Legal Aid”, a group for providing support and legal advice to socially excluded individuals.

Of particular relevance to EUNOMIA, is ABA’s Library of legal documents, the biggest in Greece with more than 80.000 readers annually. ABA has an electronic bank of a big variety of legal documents including national legislation ones, national case-law, case-law of the highest courts, harmonization documents for the integration of EU commission rights and the national law system, decisions of courts and follow up of the flow of such decisions etc. An automatic indexing and retrieval tool has been developed for facilitating lawyers from all over Greece, as well as the Ministry of Justice and other juridical and public services in accessing such information.

### **Key personnel**

**Irene Athanassiou** is a member of ABA, an advocate at the Greek Supreme Court of Justice. She has an MSc in Criminal Law and a BA in Law from the University of Athens. She has worked at various law firms on commercial, public and civic law. She has also worked at the Law Review Department of the legal library publishing house, where she used legal databases for compiling and co-authoring collections of legal cases for various publications of law-related scientific books. She is experienced in using both Greek and International legal databases.

**Eleni Gklegkle** is a member of ABA, a legal advisor of the International Organization of Migration (IOM) and of the national organizations/foundations on domestic violence, neglected and abused children and children with special needs. She has a BA in Law from the University of Athens with a specialization on family law and immigration law. She has also served as a legal advisor for various EU-funded projects (LEONARDO, LIVING CASTLE, PHARE , EUROCOALIFICATION) and for a number of Greek companies. She has delivered seminars organized by IOM, the Supreme Court of Greece, the National Centre of Public Administration and other public authorities. She has written a number of articles in magazines and on-line fora.

**Theodora Kritikopoulou** is a member of ABA, a lawyer specialized in media and human rights related issues and responsible for the monthly edition of the ABA magazine. She has an MA in the Philosophy of Law and an MA in the Sociology of Law from the University of Athens and a BA in Law from the same university. She has attended a number of seminars on property law, criminal law, commercial law, environmental law and the judicial protection in European courts. She has authored a number of articles in legal magazines and in newspapers. She is an active user of the ABA bank of legal information.

**Mary Skarsouli** is a member of ABA, a lawyer working for the Athens Bar Association. She has a BA in law and a BA in economics from the University of Athens. She has been working on the ABA online bank of legal information, and in particular on the Greek Law section; she is responsible for the online presentation of the evolution of legal clauses over time, the indexing of legal clauses according to entries in the ABA legal thesaurus and the online presentation of cross-references between clauses. She has delivered a number of seminars to lawyers and judges on effectively searching and retrieving documents from the ABA bank of legal information using the in-house retrieval engine.

### **Greek National Printing House**

The Greek National Printing House (GNPH) is the official publisher of legal acts and of the Governmental Gazette of the Hellenic Republic. The House is supervised by the Ministry of Interior, Public Administration and Decentralization and is directed by a Special Secretary appointed through a joint decision of the Prime Minister and the Minister of Interior, Public Administration and Decentralization. The organization and operation of the National Printing House is basically regulated by the provisions of the Presidential Decree 188/1996. Act 301 defines which legal acts have to be published in the official Gazette. The GNPH is responsible for the publication and circulation of the Governmental Gazette and for all printing demands of the public services. The Governmental Gazette is the only means of publication of the deeds of the State Bodies. The laws and other legal acts are published separately in different distinct issues. All Governmental Gazette Issues are available for the citizens in printed and electronic form. Some publications are free of charge while others are available upon payment of a small fee, depending on the type of the publication.

The Greek National Printing House (GNPH) has a search and retrieval engine for offering easy-access to its publications online. This system was the result of a co-operation with the Institute for Language and Speech Processing (ILSP). The system developed is part of the electronic services offered to the public by the GNPH and it offers rapid, easier and more effective search for all governmental and normative documents published by the House. The general public, legal professionals and public services have used it extensively.

#### **Key Personnel**

**Anastasia Vavatsikou** is a software engineer at GNPH where she has been working since 2001. Before that she worked for three years as a technical support engineer and as a software developer for the analysis and development of computer-based applications for the public sector. She has a BSc in Software Engineering from the Technological Institute of Athens. She was one of the main developers of the GNPH search and retrieval engine.

**George Tsipis** is a political scientist at GNPH where he has been working since 2000. He has a BA in political sciences from the Department of Law of the University of Athens. He has worked on legal document analysis and classification and legal knowledge engineering.

#### **COPIUR - Office for the Electronic Publication of Legal Data, Federal Office of Justice (Switzerland)**

Copiur, the coordination office for the electronic publication of legal data, started its work in October 1998. Initially attached to the Federal Chancellery, it later joined the Service of "Legal Data Processing and Computer Law" (Rechtsinformatik und Informatikrecht / Informatique juridique, droit et informatique / Informatica giuridica, diritto dell'informatica) at the Federal Office of Justice. The Service initiates and leads informatics and organizational projects of national importance in areas like registers, electronic exchange of legal documents and electronic publication of legislative data.

Copiur evaluates and promotes new information technologies in the legislative field. It deals principally with the elaboration of uniform norms, standards, and information structures. Its principal objective is to harmonize federal, cantonal as well as private sector publications, in order to give the public a rich, uniformly presented and simply accessible online-offer of legislative data. Copiur also represents the federal administration at the appropriate national and international coordination bodies.

#### **Key Personnel**

**Urs Paul Holenstein.** A law graduate of the University of Bern, Urs Paul Holenstein, joined the federal administration in 1992. He first worked as a scientific expert to the Secretariat General of the Department of Justice and Police. In 1998 he moved to the Federal Chancellery, where he

directed the newly-created Coordination Office for the electronic publication of legal data (Copiur). As a member of the management of e-Government projects of the Federal Chancellery, he has been closely associated with the developments concerning Federal portal ([www.ch.ch](http://www.ch.ch)) and E-voting projects.

Since 2002, Urs Paul Holenstein has been in charge of Copiur under the auspices of the Centre for Legal Data Processing and Computer Law, which is part of the Federal Office of Justice.

**Ardita Driza Maurer.** A law graduate of the University of Geneva, Ardita Driza Maurer joined Copiur in 2003. She has coordinated work with external experts, French speaking cantonal chancelleries and international experts in the field of legislative XML as well as in different projects (CHLexML, CH – Gesetzesdatenbank).

## **SWORD Technologies S.A**

SWORD Technologies S.A. (<http://www.sword-technologies.com>) is a Luxembourg- based IT company with major offices located in Brussels and in Luxembourg. The company was founded in 1999 and in 2004 it has reached the size of 240 employees and an annual turnover of €24 million. The company belongs to the SWORD Group (<http://www.sword-group.com>) of IT companies, which is listed in the Paris Stock Exchange and with major presence in France, U.K., Benelux, U.S.A and India. In 2004 the Group had an annual turnover of €79 million, a 42% increase over the 2003 figure, and currently employs over 900 IT professionals. The Group is a European leader in the area of Electronic Content Management (ECM) and Document Management. The Group's ECM solutions and services enable its customers to comply with the most stringent regulatory requirements. In the area of legal text, over the last five years it has been providing services to the European Parliament and the European Commission IT services for the consolidation and electronic publication of legal texts.

The core business activities of SWORD Technologies are the following fields:

- E-Government application development and associated IT services
- Enterprise Content Management and Document Management
- Business Intelligence and Data Warehouse applications
- System integration and tailored-made software development

SWORD Technologies provides to its clients the entire spectrum of services for application development, with its solutions being characterised as highly scalable and inter-operable, offering seamless integration with existing applications, enabling the convergence of diverse technologies, having lower maintenance costs and providing significant improvements in operational efficiency.

A notable project demonstrating the breadth and depth of the company's expertise is the area of working with legal texts is the PLAN-PUBLI project of the Publications Office of the European Union in Luxembourg, a 2 million Euro 4-year on-going project, which is concerned with the provision of an integrated IT system for the management of the publication of L, C and S series of the Official Journal of the European Union, and the provision of the underlying document management solution.

Both the company and the Group have made a strategic decision to expand their solution portfolio and the associated services both in the broad area of e-Government solutions and in the segment of electronic dissemination of legal and official information. In the next five years the Group will make a multi-million Euro investment in the development of new e-Government solutions, complementing its existing product portfolio, and seeking to reach annual sales from the new ECM solutions in the order of €4 million at the end of the five-year period. The Group has identified the area of IT services and tools for the electronic dissemination of official and legal information as an enabler to expand the appeal of its solutions to new market segments, and also enhance its offering to the existing client base.

The goals of the EUNOMIA project fit very well with the strategic objectives of the Group, which in broad terms are summarised as annual turnover growth of at least 30% and entry in new markets.

The company aims at packaging the results of the EUNOMIA project with its own e-Government and information dissemination solutions and market them to the European and world-wide market.

### Key Personnel

**Dr. Dimitrios GRITSIS** is a Senior Consultant in the R&D Department of SWORD Technologies, where one of his main tasks is the coordination of the company's activities in the R&D area; an important component of his work is the cross-fertilisation of the company's R&D activities with those of the SWORD Group.

He was from mid-1999 to the end of 2003 the International Marketing and Sales Director of EUROPEAN DYNAMICS, Athens, Greece, where he was responsible for the sales of the company whose total value over the 4.5 years period was in the order of €75 million. In his capacity, Dr. Gritsis was also the contract manager of the following IST projects: LIAISE, IST-1999-10390; EPI-SPARK, IST-1999-10392; EDEN-IW, IST-2000-29317; DIASTASIS, IST-2000-31083.

From early 1997 to mid-1999 he was a senior project manager at Intrisoft, Athens, Greece, which at that time was the largest IT company in Greece. Dr. Gritsis was the project manager and technical architect of the European Central Bank H1&H2 applications (Euro market liquidity management and foreign exchange markets on-the-spot intervention).

From 1993 to early 1997 he was the R&D Manager of Epsilon Software, Athens, Greece. During that period, he participated in numerous European R&D projects including Esprit (EP 22246 FORMULA, EP 20597 CHAMELEON, EP 6524 MIPS), Health Telematics (HC 1001 Ambulance, A2054 TELEPRIM, A2018 IREP, A2013 ISAAC), Environmental Telematics EN 1004 E-MAIL, ESSI (10632 OOSDOP), and Hellenic Esprit Special Action HYCON 7505/TR16.

Dr. Gritsis holds a Ph.D. in Systems Engineering from Imperial College, London and is a graduate of the National Technical University of Athens.

**Dr. Patrice LATINNE** is an Engineer-Consultant in SWORD Technologies. He has a degree in Civil Engineering in Electro-mechanics (specialisation in control and robotics) and a Ph.D. in Applied Sciences from the Université Libre de Bruxelles (ULB).

The last three years is involved in Data Warehouse and Business Intelligence Information Systems engineering, user requirements analysis, decision support systems analysis/design/integration, logical/physical modelling, functional analysis, prototyping and geo-marketing analyses.

From 1998 to 2002, he was a Research Fellow in the Artificial Intelligence department (IRIDIA) in ULB. At that time, he work in scientific and industrial projects (e.g. a multiple decision trees system for remote-sensing data classification – (ARC), an automatic decision support system for the real-time detection of glass defects, etc) as well as in medical research projects (e.g. data mining on astrocytic brain tumour diagnosis and prevalence using neural networks and multiple decision trees). Mr. Latinne has numerous publications in scientific journals.

**STREP Project Effort Form****Full duration of project**

(insert person-months for activities in which partners are involved)

Project acronym -

EUNOMIA	CNR	LCL	CNIPA	ILSP	IUS	RCIL	CST	RIL-HAS	IRION	ABA	GNPH	COPIUR	SWORD	TOTAL PARTNERS
---------	-----	-----	-------	------	-----	------	-----	---------	-------	-----	------	--------	-------	----------------

Research/innovation activities														
WP 2	10	9	1,5	4		3,5	2			2	1,5	2	1	36,5
WP 3	20	5,5		17	7	2,5	9	12	11	1,5	1,5			87
WP 4	20	5,5		14		4,5	10	8	11	1,5	1,5	4		80
WP 5	6	3	0,5	2	20	1	2			0,5	1,5	2	19,5	58
WP 6	12		0,5	2	10	1,5	2			0,5	1	2		31,5
WP 9	2		1,5	1		1	6			3	3	2	2	21,5
WP10	3	2	4	1	1	2	1	1	1	5	5	5	19,5	50,5
Total research/innovation	73	25	8	41	38	16	26	21	23	14	15	17	42	359

Demonstration activities														
WP 7	5		0,5	2			2		11	0,5	1,5	2	3	27,5
WP 8	5	3	0,5	2	20	4	2	0	0	0,5	1,5	2	0	40,5
Total demonstration	10	3	1	4	20	4	4	0	11	1	3	4	3	68

Consortium management activities														
WP 1	20	0,5	0,5	0,5	0,5	0,5`	0,5	0,5	0,5	0,5	0,5	0,5	0,5	26
Total consortium management	20	0,5	0,5	0,5	0,5	0,5`	0,5	0,5	0,5	0,5	0,5	0,5	0,5	26

TOTAL ACTIVITIES	103	28,5	9,5	45,5	58,5	20,5	36,5	21,5	34,5	15,5	18,5	21,5	45,5	459
------------------	-----	------	-----	------	------	------	------	------	------	------	------	------	------	-----

### **B.4.1 Sub-contracting**

#### **Schultz**

Schultz Information is Denmark's leading supplier of law documents in electronic form. Schultz will provide at least three important elements to the project: first of all expertise in presenting legal information to citizens and institutions in Denmark, secondly knowledge about Danish legal documents and their handling, - a prerequisite for Denmark's participation. Additionally, Schultz has a very wide experience in using XML as a mark-up for legal texts. Schultz will in particular participate in WP3, WP4 and WP9.

Schultz Information is a wholly owned subsidiary of The Schultz Group headquartered just outside of Copenhagen, Denmark.

The Schultz Group is owned by the J.H. Schultz Foundation. The ownership structure implies that there are no external shareholders. The aim of the foundation is exclusively to ensure that all the company's earning remain in the company for investment to the benefit of business development and the customers.

Today, the Schultz Group is one of Denmark's largest publishing houses serving professional customers with updated and value-added information in all legal fields. Thanks to a continuous technological transformation at all levels, they are now offering a wide choice of electronic products and services as well as printed publications. They are experts in cross-media publishing offering solution-oriented services via the media that best suits the customer, whether it is Internet portals, value-added online compilations of legal documents, printed publications or CD-Rom products.

We regard our self as pioneers at using XML in publishing technologies. Working with complex data structures, we constantly produce documents as well as keeping the existing data up to date, and all documents are stored in XML. Our database includes more than 150,000 XML-documents. We produce about 80,000 new A4 pages a year in XML as the database is extended continuously with new national and international laws and regulations.

#### **Key personnel**

**Mikkel Wivel** graduated as librarian from The Royal School of Library and Information Science, Denmark in 1995, and is project manager at the department for IT and Production at Schultz Information. He has 10 years experience in working with conversion of legal documents into SGML and XML format, developing Schemas/DTD's and using XML documents in publishing technologies. He was in charge of developing a MS Word plugin, eXportXML.dk, which is a tool for creating XML documents based on MS Word. The program is aimed at end-users with no XML skills, but who creates content that are needed in XML format, for example for use in media-independent publishing. He is responsible for the training in practical XML of employees at Schultz, and has contributed at seminars regarding XML and publishing.

#### **Department of Statistics, of the University of Florence**

The University of Florence is a public institution that aims at the free elaboration and transmission of knowledge in the field of higher education.

The Department of Statistics is the structure that, in the University, promotes, plans and coordinates the research in the field of statistics and its applications and that administrates resources available for those activities.

In detail, among the main activities managed from Dipartimento di Statistica are:

- coordination of teaching and administrative management of the “Corso di Laurea” in Statistics;
- management of Doctorate of Research in Applied Statistics;
- direction of a Statistics Laboratory provided with computer rooms and a calculus equipments;
- commitment in many researches and in the activities of national and international scientific organizations.

The research is a sector that keeps very busy the department personnel. The responsibility or the qualified presence in the most important sectors of Statistics Research, put the members of the Dipartimento di Statistica in a high position of the Italian statistics research. The important scientific activities worked out from Dipartimento di Statistica members is also testified from publications and participations to many national and international meetings.

The lack of adequate tools for quality control in the production of official documents represents one of the flaws in public administrative processes in particular in the planning and passing acts. The absence of quality control may give rise to serious drawbacks such as lack of transparency and readability, negative impacts on the legal system and also bad effects on society, economy or environment.

All these problems are particularly serious from the point of view of Governance as they act as barriers both between the citizen and the public administration and between public administrations themselves.

The main objective of the research is to provide the public administration with the tools needed to perform satisfactory quality control on legislative documents. The role of the Statistical Department researchers is that of assisting all the quantitative aspects of the research and particularly those directly related to the tools that will have to be advanced in order to evaluate the different dimensions of the quality of legislative documents.

In fact, in evaluating the quality of laws one cannot avoid the construction of ad hoc indexes in order to assess the quality degree of the dimensions cited above which characterize each administrative act or law. We refer mainly to indexes to evaluate the quality structure of juridical language and to measure the adherence of the text to technical standards; indexes to assess the impact on the whole legislative system (the internal impact) and survey the impact on society, economy and environment (external impact).

All those indexes will contribute to a global measure of the evaluation of what may be viewed as a statistical latent variable, i.e. the global quality.

### **Key Personnel**

**Andrea Giommi** Professor of Statistics at the University of Florence (Dept. of Statistics). From November 1997 to october 2002: President of "Corso di laurea" in Economic Statistics at the University of Florence. From august 1991 ISI member;

Scientific interests:

- Main scientific interests: survey sampling theory, with particular attention to the theory of nonsampling errors, especially for nonresponse, and to the theory of the analysis of data coming from complex surveys.

- Particular scientific interest from 1995: (a) capture-recapture methods and their application in social and economic sciences for determining the size of elusive populations (National coordinator of 40% MURST research project); (b) 'informative' sampling methods for selecting environmental data.

**Cristina Martelli** is Professor in Economic Statistics; Statistical Information Systems (University of Florence, Department of Statistics). Her research interests include Decision supporting information Systems; Statistical methods and statistical information systems for administrative processes and politics evaluation; new technologies in the production and process of statistical data (survey design, data capture, data processing, data dissemination); Statistical data bases; methods for historical and economical data base projecting and managing; Data maning; Data warehouse; Linguistic approach to statistical information sysytems conceptual modeling.

Specific experience: Consultant for Centre Européen pour la Statistique et le Développement Cesd and Unicef; in this context has developped a three years.training project for the Romanian national Statistical Institute ;

Consultant for Mozambique National Statistical Institute for the setting up of a statistical information system oriented to the study of informal economy.

Consultant for Regione Toscana and for Regione Emilia Romagna, in this context has projected and directed the developing team of an integrated information system for monitoring working conditions on the 'Florence to Bologna' High-Speed Rail Track

Consultant for the Municipality of Florence and for the Florence Province for projecting and developing an information system in support to governance Consultant for the national "Commission di Garanzia per l'informazione statistica".

Consultant for the European Project EDA Employability, Development, Adaptability VS/2002/0379: in this context an information system for the monitoring of the application of euroipean project financing laws amnong Sardinian Local Administration has been developed.

Professional Experience Record: Researches and teaching (at the "Corso di Laurea in Statistica", ); .

Publications: about 40 monographs and papers (including, international scientific journals andt Congresses).

**Emilia Rocco:** Graduated in Statistics (summa cum laude) from the University of Florence (Italy) in 1995.Visiting fellow at the Department of Statistics of the Pennsylvania State University in 1998. PhD in Applied Statistics from the University of Florence (Italy) in 1999. Employed in statistics activity at the Tuscany district from 1999 to 2001.

Researcher in Statistics at the University of Florence, since 2001. Research interests concern sample surveys and in particular informative sample designs for rare and clustered populations, such as hidden human populations and populations examined in many environmental and natural resources studies, capture–recapture methodologies, telephone and web surveys, nonresponse in sample surveys.



#### B.4.2 Other countries

**COPIUR** Coordination Office for the Electronic Publication of Legal Data Federal Office of Justice, Bern, Switzerland

(<http://www.rechtsinformation.admin.ch/copiur/index.html>) has been involved in EUNOMIA project for its wide experience in dealing with topics related to standardization of normative documents for the Swiss Confederation, and for the specific skill in dealing with multilingualism issues.

Copiur currently deals with two relevant projects.

The first one is CHLexML ([www.chlexml.ch](http://www.chlexml.ch)), or the establishment of a comprehensive XML schema to be used for the publication of all (federal, cantonal and even communal) legislative acts. The schema has been finalised by a working group of the Swiss association for juridical informatics led by Copiur and the Federal Chancellery. After an internal review, it will be adopted at the 5<sup>th</sup> seminar on law and informatics organised by Copiur on 30 June – 1<sup>st</sup> July 2005. The schema CHLexML will then be submitted to cantonal chancelleries and other interested players for consultation and to the eCH (the E-Government standards setting association) for normalisation.

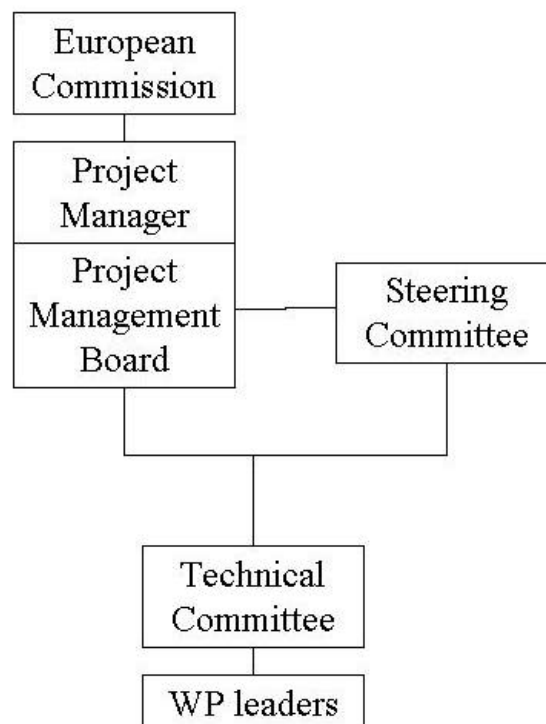
The second project is named LexGo ([www.lexgo.ch](http://www.lexgo.ch)). Its main purpose is to align the 27 classification systems (all different) used by the 27 (federal and cantonal) systematic collections of law in order to simplify and improve search results. Based on the common systematic for the classification of legal acts elaborated by the Institute of Federalism of Fribourg, LexGo has created 27 matrix tables (Konkordanztabellen, tableaux de concordance) aligning all cantonal and federal legislative acts to the common systematic. Thanks also to a database of accurate links, LexGo allows users to find, for example, all federal and cantonal norms related to a given subject.

COPIUR will play a central role in evaluating the effectiveness of the European legal standards proposed by the project. Moreover, taking into account the multilingual legal scenario it used to face, it will be the main reference to test multilingual services built on enriched legal data produced in different cantonal languages.

## B.5 Project management

The procedures for project management within the project will be detailed in the project's "Project Management Plan" which is due at M2 from the commencement date specified in the contract with the EC. A consortium agreement shall provide rules and terms of reference for any issue of legal nature concerning the co-operation among the parties as well as the intellectual property rights of individual partners and the consortium as a whole.

### Organisation and management structure



The EUNOMIA project will be coordinated and run, within the activities of WP1, by the Project Management Board (PMB) which will be formed by the leading staff of the two institutes (ITTIG and ILC) of the Italian National Research Council (CNR), the EUNOMIA coordinating partner. The PBM will be responsible for the overall management of the project as will be specified in the contract with the EC. For this purpose a project manager will be assigned, who will be responsible for day-to-day project coordination issues.

### Steering Committee

Each partner will appoint one person to be part of the Steering Committee (SC). The SC will have the overall responsibility for the technical, financial, administrative and exploitation and dissemination aspects of the project. In particular, the SC will:

- decide the strategy for conducting the project according to the terms of the contract

- assess the progress of the project, adopt any corrective measure if necessary and authorize appropriate amendments to the work plan according to the recommendations of the Technical Committee in order to meet the project objectives
- review the policy and strategy for exploitation and publicity, authorize the Project Exploitation Plan and any revision should become necessary
- assess the impact of any change to the contract suggested by the EC and respond accordingly
- resolve any conflict, technical, managerial or financial that may appear amongst the project team members during the project duration.

The SC will normally meet on a quarterly basis and as often as required.

### **Technical Committee**

The technical management of the EUNOMIA project will be performed by the Technical Committee (TC). Members of the TC are the leaders of the technical workpackages (WP2 to WP10).

The role of the TC is to determine the technical direction of the project in order to fulfill its aims and those of the consortium members. In particular, the TC will:

- Ensure the attainment of a consensus on the technical content of the project as a whole as well as at the detailed levels
- recommend justified technical amendments to the project work programme
- perform a quality control of the work carried out during the project.
- maintaining regular e-mail communication and problem-solving dialogue with all partners to ensure that they are regularly updated on progress;
- reviewing achievement of project goals against updated time plan, for circulation on a regular basis and signalling any need for remedial action to the SC;
- proposing and implementing evaluation mechanisms (usage, satisfaction, impact assessment etc) for each aspect of the work;
- reviewing all draft and final deliverables and making arrangements for independent review where appropriate.

The TC will normally meet on a quarterly basis and as often as required.

### **WP leaders**

Each WP falls under the responsibility of a consortium partner (WP Leader) and within the workpackages Task Managers responsible for a set of specific activities will be addressed. They will report activities to the workpackage leaders.

On top of the clear allocation of responsibilities and quality control, strong attention will be paid to regular day-to-day communication between the partners. This will be conducted by e-mail (a news list will be set up) and telephone, actively facilitated by the Project Management Board. Furthermore, a website will be established further ensuring effective communication on the project with partners and target groups sought (publishers, potential end users). Furthermore, contacts will be established with relevant European and national projects to exchange knowledge allowing for cross-fertilization.

**Conflict resolution and risk management**

Should any conflict, either technical, managerial or financial, arise amongst the project team members, the difference will be resolved by the SC. Detailed dispositions will be precisely defined in the Project management plan and endorsed by the project representative of the SC by the end of M3. These dispositions will be in line with the current recommendations of the EC policy and the standard consortium agreement.

**Quality control and risk assessment**

In the EUNOMIA project, the Project Director will be in charge of overseeing the work, results and quality of the technical work produced in mainly workpackages 2 to 7. Each workpackage, task and project result will be under the responsibility of a single partner as indicated in the description tables in Section B6. Workpackages are further split into tasks to assign responsibilities to well identified people and to get regular milestones to measure project progress against plans.

Co-ordination meetings will be held regularly (at least every 4 months and more often if required) to co-ordinate activities and to ensure the consistency of activities in respect with the schedule.

Some specific tasks will receive a particular attention:

- Self sustainability, i.e. validating the economic model and prepare continuation outside EC support
- Information Dissemination, i.e. widening the scope of the EUNOMIA industrial user base (Europe and international) aware of EUNOMIA and support EC communication efforts when required

In order to address these aspects efficiently, the partners will either appoint experts with the required level of experience from inside their own organisations or through outsourcing. Supervision of these tasks will be handled directly by the Project Board.

Effective dissemination will be measured by the number of visits to the EUNOMIA web site, the number of downloads of the EUNOMIA prototypes and software and the number of participating industry Advisory group members.

In addition to the high level risks already identified, the following contingency strategy matrix has been put in place:

<b>Risk</b>	<b>Probability</b>	<b>Impact</b>	<b>Contingency Strategy</b>
Partner leaves consortium	Low	Medium	Consortium is of sufficient strength and diversity for other partners to replace if required.
Staffing & Recruitment Problems	Medium	Low	EUNOMIA has a strategy of technical excellence in given disciplines spread across the partners which can be seen as a risk balancing to avoid excessive dependency on any one partner.
Key staff illness during critical project phase	Medium	Medium	Critical parts of project executed by more than one partner.

Technology changes require redesign	Medium	Medium	The Project Board will monitor external developments and use them to re-plan if necessary.
Time for development is underestimated	Medium	Low	Project checkpoints will monitor, detect problems early and take corrective action.
EUNOMIA standards are not accepted by the users	Low	High	Key efforts are being set up to define an industry driven strategy for standards development (see Industry Advisory Board). The EUNOMIA consortium already includes the EU-leading sites carrying out research on legal modelling, so take-up by the community will be ensured through wide dissemination via publications and web sites. There will be several workshops aimed at ensuring take-up.

There are a number of critical relationships between tasks and deliverables in the project. These are described and shown in the WP tables in section B6. Project progress will be closely monitored and reported formally to the Commission on a six monthly basis. There will also be more frequent informal contacts between the coordinator and the Commission, as required by either party. In this way any problems will be identified at an early stage and remedial action will be taken and reported to the Commission.

## **B.6 Detailed implementation plan**

### **Introduction**

To achieve its objectives the project is structured in a number of activities organized in ten strongly interlinked workpackages (WPs). WPs have been shaped around the mission critical activities of bringing the project forward, towards a successful definition of European standards for legal documents, at the exploitation of its potentialities regarding the development of tools to make the adoption of such standards easier and to develop high quality services for citizens and law makers.

Activities are organized according to a gradualness principle so that each WP has clear and reachable objectives. Goals and end-points of each WP will have the form of deliverables; on their basis, the activities of each WP will be monitored and evaluated by the project management.

### **WP1: Management and coordination**

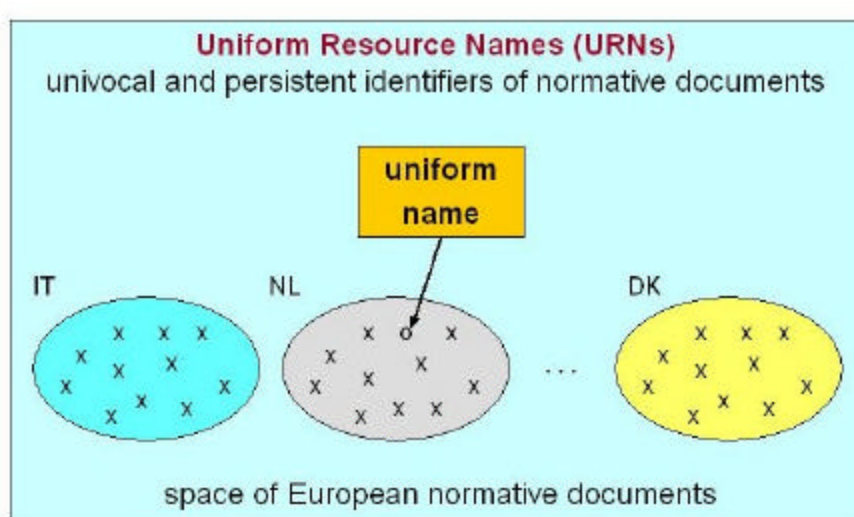
Within WP1 management and coordination activities of the whole project will be carried on, in order to ensure a smooth project organization and quality control, as well as clear decision making procedures. Activities of WP1 will be developed throughout the project.

Particularly, at the beginning of WP1 a detailed workplan will be established, based on the description in this proposal which establishes to break the various WPs down into specific tasks. Nevertheless, before submission of this proposal agreement has been reached on the contours of the work to be done and each WP has a single partner clearly responsible for the completion of the deliverables of that WP (as indicated in the Workpackage overview).

### **WP2: Standardized representation of legal documents**

As a preliminary activity of this WP, an analysis of the scenario of normative document production in different European countries will be carried out, in order to assess potential barriers to the harmonisation of the models for normative documents representation. Taking into account this survey, and the experiences on legal standards at national level (for example NIR project in Italy, LexDania in Denmark, MetaLex in Netherlands, CHLexML in Switzerland) the following activities of WP2 will be devoted to the definition of a core normative document representation model including facilities to formalize the most relevant structures and metadata of European normative documents, factoring the individual differences and unifying commonalities. These facilities will deliver at least three different formalisms for the identification and description of normative documents:

- A formalism for normative document identification, based on URI/URN techniques, allowing to univocally identify any normative measure issued by any European authority (Fig.3). It will provide also a way to express references in a reliable and long term way independently of the physical locations of the actual documents;



**Fig. 3 Space of European normative documents and URNs**

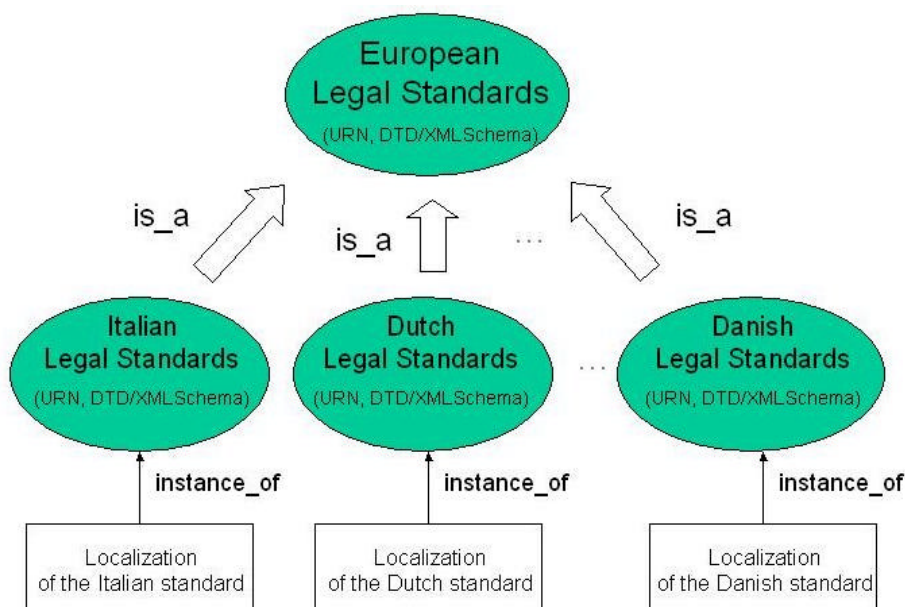
- A formalism for the representation of the actual structure of normative documents, based on XML and issued in one or more of the existing schema languages (DTDs, XML Schemas, Relax NG, etc.), describing the meaningful structural partitions in normative measures of different countries.
- A formalism for the description of the content of normative documents, at two levels of detail:
  - General: general information about the normative act contained in the document, such as subject classification, publication date, issuing body or organisation, relationships among acts.
  - Analytic: finer-grained information about the semantics of the normative documents, where by semantics we mean the particular provisions expressed (i.e. an obligation, a permission, a penalty, etc.).

To this end, it is expected that a model organised into classes or types of provisions will be adopted, especially in the view of different levels of depth at which a normative document can be described, as far as its content is concerned. In addition to the provision type, a semantic description of normative documents must also make explicit the conceptual roles that are relevant for any particular type of provision. For instance, there are at least two roles that are relevant for the description of an obligation: whom the obligation is addressed to, what the obligated person/entity is obligated to do. Similarly, if a text expresses a modification to a previous text such as a repeal, it is interesting to extract which is the text to be amended, which part, and which particular portion of it is affected by the modification.

The project proposes to adopt XML and RDF as common languages to represent in a standardised way both the formal and the semantic structures of normative documents at the European level and the URN standard to univocally identify them on the Web.

The standards for European law documents will be configured as the minimum common framework with which law documents of different European countries can be described.

This standard should be the result of a generalization of the standards already established by some European countries; similarly it should be a reference for those countries where standards for law documents have not yet defined, so to obtain local new standards as specializations of the European one.



**Fig. 4 Relationships among the European legal standards proposed by the project and the national legal standards.**

Further than the extensions of the European standards in terms of DTDs or XML-Schema, the different national standards could be mapped to the same specification expressed in the local language, expressed in terms of DTDs or XML-Schema as well.

### **WP3: Design of automatisms for legal document annotation**

This WP is concerned with the design and implementation of mechanisms for the annotation of normative documents, at the three levels described above.

WP3 logically presupposes WP2, since the automatic annotation will be performed according to the specifications set out in WP2. WP3's role in the project is significant for two reasons:

- a) it facilitates and promotes the adoption of the common representation model defined in WP2 by offering the technology needed for automating the application of the suggested legal standards on normative documents.
- b) It permits the development of high quality services needed by both citizens and legal practitioners (as WP4 also does); the services developed in WP5 (legal text editing), WP6 (text quality control), WP7 (multilingual and cross-lingual indexing and retrieval) and WP8 (text reconstruction) rely on the existence of the annotation mechanisms of this work package, without which they would not be able to offer "intelligent", advanced features that go beyond state of the art document management prototypes (cf. sections B.1 and B.3).

The availability of automatic tools for the annotation of normative documents is particularly relevant for the task of semantic annotation of normative documents. We remind that by the term "semantic" we refer to the twofold task of i) assigning a portion of a normative text to a provision type and ii) rendering the conceptual roles relevant for a particular type of provision explicit. This activity can be seen as an intellectual activity, which is part of the everyday tasks of law makers, legal experts and practitioners. However, it can be particularly time consuming, especially for long and complex laws.



Moreover, automated, standardized and authoritative classification of norms can be of great help to citizens wanting to access normative documents even though lacking the necessary expertise.

The work inside this WP will be organized into two main tasks each with its own subtasks:

### *T.1 Automated tools for unique identification and structural annotation of normative documents*

To obtain the automatic transformation of legacy legal documents into their XML representation, a number of parsers will be implemented able to annotate the structural partitions of a law text as well as to detect and mark the cross-references contained and to construct the related URNs.

A law text may contain lots of cross-references to other measures that have to be described using the related URN, so that references can be transformed in effective links when documents are published on the Web. Information to build URNs are usually contained in the citation (for example the citation: “Act 24 November 1999, No. 468” generates the following URN-NIR “urn:nir:stato:legge:1999-11-24;468”).

Especially in the phase of legacy content conversion, the manual construction of a URN for each reference can be a time-consuming work. For this reason a module able to automatically parse a law document, detecting cross-references and assigning them the related URN will be developed. Such a module will have to know the nationality of the act a cross-reference is related to, then it will parse the sentence expressing the reference, mark it as a reference and automatically construct the related URN.

Similarly, a structure parser able to transform a legacy law document into an XML format according to the standards defined in WP2 will be developed. The expected document native formats will be HTML and plain text.

### *T.2 Automated tools for semantic annotation of normative documents*

The aim of this subtask is the development of tools which take as input portions of normative texts and output an annotation of the same text with tags whose purpose is to make explicit: a) what kind of provision is expressed by a certain paragraph (an obligation, a penalty, a permission, etc.); b) which portions of the text correspond to the conceptual roles that are relevant for any given type of provision, according to the model developed in WP2. For instance, given a text such as “The Contractor shall notify the Commission twenty days in advance”, the automatic semantic annotation procedure shall be able to render us an annotation along the following lines:

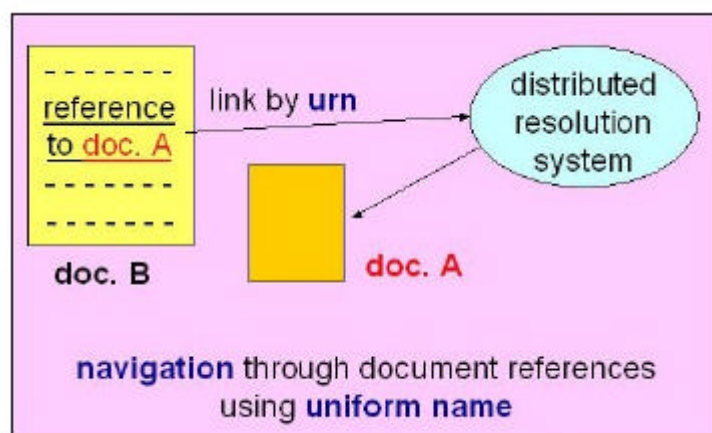
```
<obligation>
  <action>
    <obligated_entity> The Contractor </obligated_entity>
    must notify
    <recipient_of_obligation> the Commission </recipient_of_obligation>
    <time_of_obligation> twenty days in advance </time_of_obligation>
  </action>
</obligation>
```

The twofold task of classifying law paragraphs according to their regulatory content, and of extracting relevant text fragments corresponding to specific semantic roles that are relevant for the different types of regulatory content will be performed combining knowledge extraction methodologies, based on machine learning, with symbolic approaches to Natural Language Processing, relying on specific language patterns that provide the necessary clues for translating a natural language text into a formal

representation. The feasibility of this approach has already been tested in (Bartolini et al. 2004; Biagioli et al. 2005), where a case study performed on a representative set of provision types yielded promising results. Being partially language-independent, the overall architecture is expected to be portable across languages.

#### *T3 Development of a prototype of a distributed resolution system*

Strictly connected to the definition of the URN standard, in order to make effective the link constructed by the cross-reference parser, a distributed resolution system will be developed. Such a service will associate a URN to one or more than one URLs (physical location on the Web) so that documents referred to by a link can be reached in a distributed environment (Fig.5)



**Fig. 5 The URN resolution system scheme**

### **WP4: Conceptual indexing of normative documents**

The principal aim of WP4 is the automatic and semi-automatic identification of the conceptual entities addresses/evoked by the normative documents for the twofold task of:

- using normative documents as a term repository which can be exploited to populate domain-specific ontologies with data belonging to real texts, thus reducing the costly process of manual ontology building;
- producing a conceptual indexing of documents with semantic metadata corresponding to the terms, concepts and relations of the ontologies.

WP4 is similar to WP3 in that it aims at producing a representation of the content and an indexing of normative documents. However, while the purpose of WP3 is to produce an indexing of documents according to the *regulatory content* of each provision, the purpose of WP4 is the identification of the *regulated content*, through the concepts, entities that play a special role within the selected provisions.

Although in principle independent, WP3 and WP4 are closely related. In particular, WP4 is intended to capitalise on the results of the semantic annotation performed in WP3, in order to be able to produce a text that is indexed with reference to the legal entities involved *together with* their regulatory context (e.g. the duties of an employer, the penalties for an employee performing a certain action). This richer semantic indexing of documents is bound to boost data quality, since it can enable sophisticated document search and retrieval applications. The following sub-tasks can be envisaged:

#### *Import of multilingual lexicons and ontologies for Law*

Lexical repositories and Core Ontologies constitute a *descriptive vocabulary* for modelling specific legal domains. The project will privilege the re-use and integration of the proposals in the field of multilingual lexicons for law and of legal ontologies, for instance multilingual localisation of legal WordNets.

The WordNet-like lexicons can be considered as standard in the paradigm of computational lexicons. Legal WordNet acts as a support in legal information searching, as leaf concepts are structured according to taxonomy and semantic relationships based on linguistic rules. In order to remove terminological and conceptual ambiguities and to face multilingual information handling, the high level concepts in the lexicon need to be framed and organised via a *Core Legal Ontology* (CLO).

#### *Term extraction and conceptual indexing*

Document indexing implies the categorization of documents according to a predefined set of keywords or concepts. Again, this type of task is often performed manually, to ensure accuracy. However, automatic tools to support this task are highly desirable.

The project intends to address conceptual indexing of documents as opposed to keyword indexing. Conceptual indexing means categorizing documents according to the concepts expressed, as opposed to the actual words used.

The benefits of conceptual indexing of legal documents are especially evident from the point of view of its implications for smart search and retrieval applications. Conceptually indexed documents can allow users to:

- Perform specific queries: for instance, a user interested in documents about "restructuring" can reduce the noise that would be normally introduced by a keyword-based query by specifying the particular concept he/she is interested in (for instance, as a type of "renewal" as opposed to a type of "rescheduling");
- Extend the query to the superordinate or subordinate concepts: for instance, a search related to the concept of "research grant" can be automatically extended to its superordinate concept "financing".
- Extend the query to synonyms: a query concerning for instance "photovoltaic roof" can be automatically expanded to retrieve documents concerning "solar panel".

Conceptual indexing of documents presupposes a suite of machine-learning and natural language processing techniques able to:

- extracting the terms contained in the legal texts and corresponding to relevant concepts;
- performing a semantic interpretation of the terms (concepts) in order to determine their right sense;
- taking into account possible definitions of the terms in the act processed or in previous acts of the same domain;
- mapping terms onto existing domain-specific ontologies;
- identifying the semantic relations holding among the concepts.

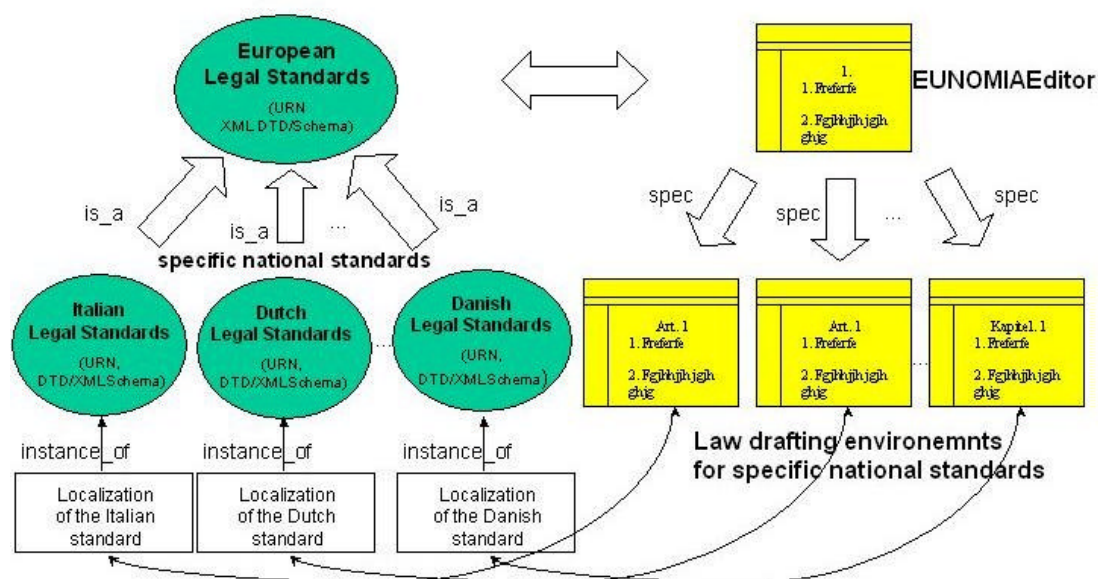
Again, the adoption of hybrid architectures, combining statistical and machine learning algorithms with NLP tools, ensures better customisability to other languages.

#### **WP5: Integrated workbench for the editing of normative texts**

In order to make easier the adoption of the established European legal standards, an editor operating within the URN and XML-DTDs or XML-Schema framework established by the project will be developed.

Such an editor, far from being a general-purpose XML editor, it will be configured as an environment specifically designed to deal with the class of documents identified and generated according to the established XML European standards.

In fact even though programs for XML drafting already exist, they have limits whether used for a specific class of documents, especially as concerns the generality and inadequacy of their editing functions to cope with the constraints established by the European law document standards. The characteristics of the editor, specifically devoted to manage a class of documents, guarantees also the feasibility of the project. Following the hierarchy of the legal standards at the European level that this project is aimed to obtain, the editor will be configured as reference for those countries where legal standards will be implemented as specialization of the European ones. Similarly editors for specific national legal standards could be implemented as specialization of the European one, according to an object oriented programming scheme (Fig. 6)



**Fig. 6 Relations among legal standards and law drafting environment working on these standards**

Further than the extensions of the European standards provided by the different national standards the editors in different countries will have to deal with the locale language extension of the standards.

The European editor will have to be able to work in two situations: it will be designed to process legacy law texts, as well as to assist the drafting of new texts. In both these two working situations the editor will handle the formal as well as the functional profile of a law text, using both manual and automatic facilities.

The semantics expressed through analytical metadata will play a central role also in the drafting process of new acts, as an help while expressing provision texts and as a guide in organising the text itself, to improve the technical quality of normative documents as expected in the objectives of the project.

Document versioning facilities and concurrent editing will be implemented as well: the editor should allow "Checking-Out" of a whole document or just part of the document, thus leaving document "locked" for the current user. After changing the content of part of the document (or adding new part), this part will be stored as new

version of that part of the document, so the whole document now exists in two versions: version prior the modification and version containing modifications.

#### **WP6: Methods and tools for the quality control of normative texts**

The aim of this WP is to provide instruments and methodologies for evaluating the quality of normative texts. The work is organised around the following tasks:

T1. the project will survey the evaluation check lists used by UE countries for legislative texts. We will verify the Italian regulation about evaluation system (see “Direttiva del Presidente del Consiglio dei Ministri del 27 marzo 2000”) and we will check if the Italian subdivision in “Analisi tecnico normative” that we can translate Normative impact and “Analisi d’Impatto della regolamentazione” (in English Regulatory Impact Analysis (RIA)) exist in other European countries.

About normative Impact we will analyse the assessment parameters who will give significant indications about the compatibility between new legislative acts and the law enforced.

About the regulatory impact analysis we will survey methodologies and criteria for various evaluations as:

administrative impact as suitability of the P.A. to reach the goals aimed by legislator;

economic impact as costs benefits evaluation;

social impact as relevance of new legislation for social relationship;

environmental impact as check of the environmental modification produced by new legislation.

T2. Based on the survey the project will propose a check list that could be a common platform for the European countries to adopt law quality standard indicators.

T3. NLP methodologies will be adopted to extract information from legislative texts, which are relevant for quality evaluations.

Data will be extracted the from the normative corpora annotated on the base on the methodologies developed in the previous workpackages of the proposal.

T4. Methodologies will also be developed to match the information extracted from normative texts with those extracted from a factual database. The factual database will be structured in the way to facilitate this matching and they will contain data representing events and states those will be altered by the new legislation.

T5. A prototype will be tested on a restricted corpus of laws (representing legislation of two countries participant to the project), annotated on based on methodologies defined in the previous WP, the normative impact and the RIA evaluations in an automatic or semi automatic way.

The normative impact assessment will be possible thanks to the connection and the match between the annotations in the text on analysis and the annotations in the legislative texts of the corpus.

This match will permit to verify the conformity to the standard defined on the check list.

The RIA will develop on the base of the methodologies indicated above and will be experimented only for the data of a prototypal database about public administration structure and organisation.

T6. On the basis of the methodologies and the techniques indicated above, law quality indices will be obtained. These quality indices could be gained with specific tools for the assessment through partial and total scores.

#### **WP7: Multilingual semantic search and retrieval of normative documents**

The purpose of this WP is to demonstrate the usefulness and usability of semantically annotated normative data through the development of an advanced search and retrieval application. The retrieval system will combine meta search with conceptual indexes on free text. The meta-search will be applied to the automatically extracted meta-information from the legal documents. The conceptual indexes will be built with wordnets that are tuned to the legal domain. The search will be cross-lingual covering all the project languages. Finally, the retrieval system will show activation of inferences for the retrieval of implicit and/or related norms (logical and legislative inferences).

The demonstration will measure the productivity of the added value for the legal domain. It will include:

- a precision/recall benchmark with a substantial number of queries derived from the corpus and translated to all the project languages
- a end-user evaluation by a number of end-users, based on pre-designed retrieval sceneries. The end-user behaviour is logged and used for measuring effectiveness.

The demonstration will use the TwentyOne retrieval system developed by Irion Technologies.

#### **WP8: Automatic reconstruction of the normative text in force**

The construction of the normative text in force in consequence of amendments is one of the main issues of the juridical-legislative system, aimed at the consolidation of the normative in force. Currently the production of the in force text is carried out mainly by publishers through an editorial activity.

On the basis of the established standards the project aims at defining effective methodologies able to obtain the automatic reconstruction of the normative in force. Such methodologies will be based on a specific descriptions of the amendment actions in terms of *amendment analytical provisions*, basically a sub-class of the analytical metadata the project will propose within WP2.

Such a semantic description can be obtained automatically (see WP3) or manually.

The combination of an XML representation of normative documents with a semantic description of the amendment actions in term of *amendment analytical provisions* will allow to develop tools able to automatically implement substitution, insertion and repeal actions.

Two main types of amendment actions can be considered:

- 1) actions on XML formal partitions (articles, paragraphs, etc.);
- 2) actions on sub-partitions (amendment action on text strings).

The project aims at implementing a demonstrator able to construct the text in force as, whose semantic description is manually or, only for the first type, automatically obtained.

#### **WP9: Evaluation**

As the main objectives of the project is to define standards for legal documents and developing tools to adopt and use these standards in an effective way, it is extremely important to evaluate constantly the proposed solutions and giving feedback to the provisional results in order to improve them, starting from the earliest phases of the project. The evaluation will be concerned with functionality evaluation as well as usability evaluation. In all cases, ISO standards for the evaluation, quality assurance

and usability (quality in use) for software development (ISO 9126, ISO 8402, ISO 9241) will be used wherever applicable. At the same time quality standards for legal documents will be used to provide the professional legal basis for the evaluation.

In work package 9 we will perform these tests in relation to the definition of standardized representation of legal documents (WP2), the design of automatisms for legacy content handling and detection and extraction of norms (WP3), and legal knowledge conceptualisation and semantic indexing of legal texts (WP4). Furthermore utility and functionality tests will be performed on the prototypes developed in WP7 and WP8. An important synergy with the evaluation work in WP6 is expected.

In the beginning of the project an *Advisory Panel* (WP10) will be created in order to give input and advice on formalisms, annotation schemes, platforms, functionalities and usability.

A central role within the *Advisory Panel* will be played by the *Official Journal of the European Union* (OJ). OJ is the only periodical published every working day in all 20 official languages of the EU. It consists of two related series: L for legislation (regulations, directives, decisions, recommendations, opinions) and C for information and notices) and a supplement (S for public procurement).

In addition, the *Directory of Community legislation in force* is published as part of the OJ L series. This directory lists references to the initial texts and to any subsequent amendments. It also includes references to agreements made and conventions signed by the European Union in the framework of external relations, binding acts under the EU Treaties, complementary acts, such as those of the Council of Ministers and Heads of State or Government, and other non-binding acts which are relevant for the institutions.

Documents published within the L series, belonging to the normative domain the project deals with, will be used to test the standards established by the project, as well as the tools to promote their adoption, and the services on enriched data.

Moreover *Advisory Panel* will comprise industrial users (SMEs), professional associations, dedicated publishers, PA representatives.

Taking into account the input of the Advisory Panel and the WP2 specification a preliminary list of evaluation criteria will be defined. In the early phase of the project a test corpus of legal texts will be collected, together with a development corpus to be used in WP3 and WP4.

The evaluation criteria defined in the first phase of the work package will be refined during the later project phases taking into account i.e. the feedback from the Advisory Panel and relevant feedback collected from the dissemination campaign (WP10) on both the standard specifications and the demonstration applications.

The main activities of the work package are the following:

- discuss evaluation methodology, define preliminary evaluation criteria for standards of structural and semantic annotation of legal documents on the basis of the WP2 specification and Advisory Panel's input
- collection of test corpus and of demonstration corpus
- define functionality and usability tests of standards of structural and semantic annotation of legal documents outside and inside the demonstration applications on the basis of the WP2 specification and Advisory Panel's input
- continuous evaluation of results from WP2, WP3, WP4, WP7 and WP8

using the defined tests and refinement of evaluation criteria on the basis of collected feedback

- final evaluation

It is foreseen that the Advisory Panel will convene 4 times during the project, at stages where its input will be most beneficial, being:

- a) at the beginning of the project in WP1, allowing the consortium to fine tune the project plan and to set the right accents in relation to the user requirements;
- b) during WP2, allowing to provide the Project Management Board with the user perspective in conjunction with the system requirements;
- c) after finishing WP4 and at the end of WP7 and WP8 whereby the Advisory Panel will test and validate the service and its infrastructure.

### **WP10: Information dissemination and exploitation**

Disseminating legal and other government information from any EU member to the public aptly expresses the foundational principle regarding public access to the national and European law. Moreover, an effective democracy is recognised that it requires ready public access to law. Yet, ensuring ready and effective public access to the primary sources of law (legislation, reports of judicial decisions, administrative agency regulations and rulings, etc.), has not been accomplished easily even in the print information environment. The transformation of much published legal information into electronic formats provides new opportunities for improving access to the law, but also raises a new set of issues for access and delivery of legal information.

EUNOMIA aims at facilitating the access and delivery of legal information to the public. The enlargement of access possibilities to the European base of legal documents is expected to have a substantial impact in the development of digital content and vice versa. The consortium will attempt to capitalise the multiple effects of this impact by providing a solid basis and extending the participation to public institutions, governmental bodies, law dissemination agencies and e-publishers of countries, which are not currently involved in the project.

The consolidation of legal information, the quality of standardised legal texts and documents generated within different legal systems, the dissemination of reliable, updated and easily accessible public legal information are some of the advantages deriving from the project. The utmost benefit is for the user of legal documents and citizen, who will be granted an advanced access level in legal information, either directly through on-line systems in the European environment and at national level, or indirectly, through the dissemination of accurate and complete legal information from professional (official or commercial) legal institutions, agencies and publishing companies responsible for serving the citizen in law related matters.

The exploitation of the technological advantages and social benefits of the tools to be developed and resources devoted for the project is one of the primary targets of the consortium and we can guarantee that a substantial part of the effort and budget will be devoted to that end, and in parallel with other actions of marketing and distributing the project results.

Foreseen marketing and business dissemination actions will be directed towards more relevant consumer and commercial categories represented by:



1. **Legal Information publishers** and industries which are technical and standards developers at European level contributing to the potential exploitation of public information, as well as the potential exploiters themselves. In fact, developments in semantic-web/web services and related standards enable the reuse and commercial exploitation of legal information and of all Public sector information as well.
2. **On-line consumers** of legal information, including both citizens and law firms: As for the citizen as on line consumers of legal information there are no estimates figures about the size of the legal sector compared to the general e-commerce sector and legal studios. Such a figure can be induced by considering that legal information sites (2343) listed by *Google* directories are 10% of the all sites devoted to programming language and roughly 70% of all on-line newspaper.
3. **Public administration**: Legal Information is part of the Public Sector Information (PSI) supply chain: agents involved are public sector organisations, re-users of PSI and the end users of the information. Within the entire process, Public Administrations play the roles of providers, users and re-users in the same context. This explains why the standardisation process in the area of PSI is driven by public policies (e.g. Interoperability Frameworks such as in France and the UK (e-GIF, e-Government Interoperability Framework adopted by Canada and New Zealand and MIREG on a EU level) more than by general developments in information technology. Project results will offer to Public Administrators a tool to Add Value in Information Delivery services, in Legal Information distribution, selling, re-using.
4. **Commercial companies**: The collection of a significant amount of information in digital media will boost the commercial exploitation of raw legal material generated by the public sector, national governments and European bodies. Private companies exploit or will exploit this raw material for producing added-value products (law documents combined from various national sources, European and national case-law on specific issues, etc) in the market of legal information. The types of added-value products include: aggregation of legal information at cross-border level, new delivery mechanisms (e.g. mobile phones), integration in a broader package of services (with multilingual characteristics, etc), tailoring of information to the needs of specific users, etc. In many cases, public sector information from one specific source or sector can be combined with other information coming from private or public sources. Examples of such similar exploitation of raw law material and information can be found in the United States with the Lexis-Nexis and Westlaw companies.

As for the results the project will exploit, they can be classified in:

### **1. Resources Exploitation**

Repositories of multilingual semantic metadata: For organisations involved in exploitation, it is of the highest importance that the creators of information apply adequate mechanisms for identification and description of the information. Main issues here are the assignment of standard identifiers (ISBN, ISSN, DOI, maybe URNs) and standard metadata (Dublin Core, ONIX or ebXML). Proposals in the field of encoding schemes and content description and management (data representation,

controlled vocabularies such as GCL and Eurovoc) are still few and not domain specialised.

## ***2. Software Exploitation***

Automatic classification, indexing and Searching strategies: Publishing and presentation strategies are driven for a large part by technical developments. Harvesting techniques and interoperability aspects can play an essential role.

Cross-lingual searching engine Web search engines are an important input in the structuring of access, combining broad-brush searching strategies with user-oriented classification and metadata search facilities. They offer opportunities to bring together collections from various sources to provide value-added services.

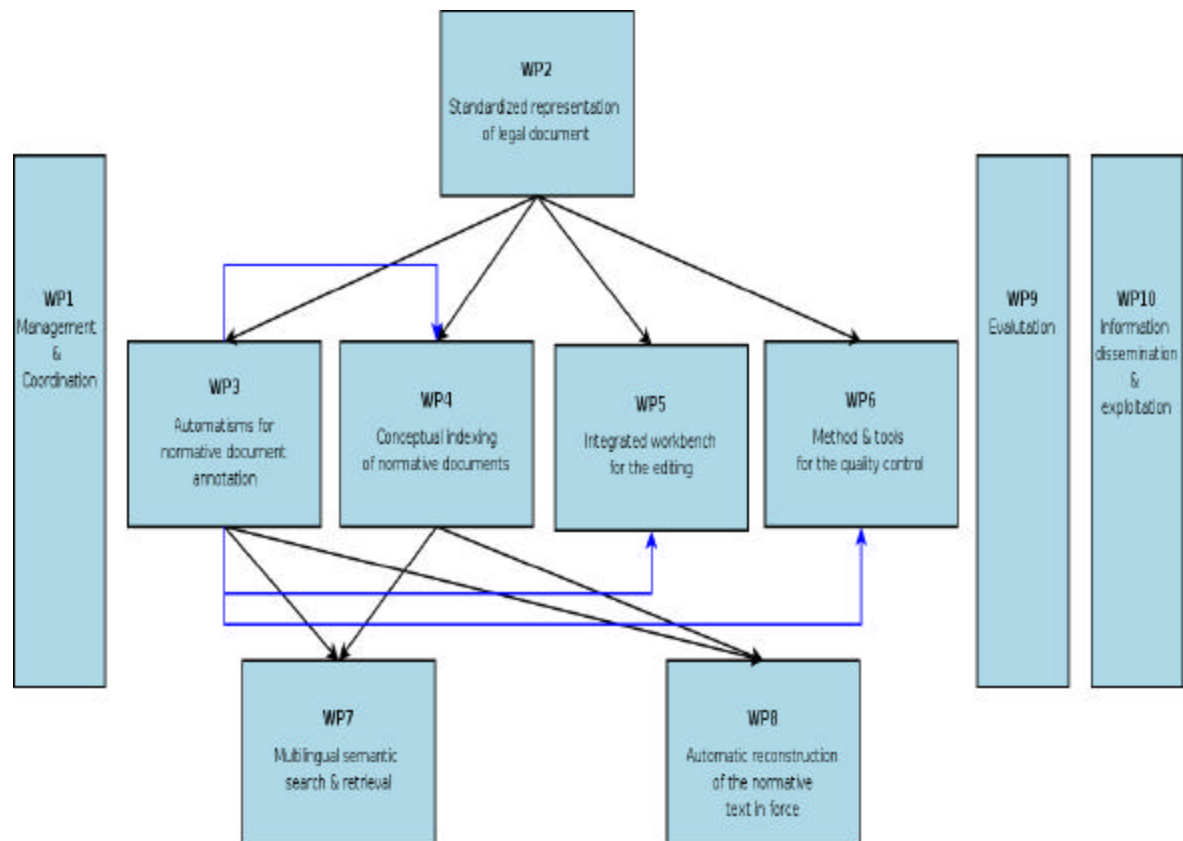
## ***3. Services Exploitation***

Standard generation: All European member countries, and the EU itself, are committed to the development and implementation of standards. Unless providers of Public Service Information understand and adhere to these standards, the information produced will not be interoperable with other sources of information and the possibilities of re-use will be severely limited. Standards add value to and reduce the cost of collecting, aggregating and providing PSI, thereby providing best value to the taxpayers who fund it.

b) Work planning, showing the timing of the different WPs and their tasks

Name	Work	2006		2007		2008	
		S1	S2	S1	S2	S1	S2
WP 1 - Management & coordination	1066d						
<b>WP 2 - Standardized representation of legal documents</b>	<b>730d</b>						
T1a - Review existing mechanisms for identification & structural description	56d						
T2.1 - Specifications of a core standard for identification & structural description (v. 1)	122d						
T2.2 - Extended standard for identification & structural description (vers. 1)	184d						
T1b - Review existing mechanisms for semantic description	56d						
T3.1a - Specifications of a core standard for general semantic metadata (vers. 1)	122d						
T3.2a - Specifications of a core standard for analytical semantic metadata (vers. 1)	122d						
T3.1b - Specifications of an extended standard for general semantic metadata (vers. 1)	184d						
T3.2b - Specifications of an extended standard for analytical semantic metadata (vers. 1)	184d						
T4.1 - Improvements of standard for identification & structural description (v. 2)	365d						
T4.2 - Improvements of standard for general semantic metadata (vers. 2)	365d						
T4.3 - Improvements of standard for analytical semantic metadata (vers. 2)	365d						
<b>WP 3 - Automatism for normative document annotation</b>	<b>549d</b>						
T1.1 - Review and identification of appropriate technologies for parsers development	62d						
T1.2 - Definition of a common set of normative documents for references & structure parsers	62d						
T1.3a - Development of a language-independent prototype of references parser engine	133d						
T1.3b - Development of a language-independent prototype of structure parser engine	133d						
T1.4a - Development of language-dependent prototypes of references parser (vers. 1)	150d						
T1.4b - Improvement of language-dependent prototypes of references parser (vers. 2)	184d						
T1.4c - Development of a language-dependent prototypes of structure parser (vers. 1)	150d						
T1.4d - Improvement of language-dependent prototypes of structure parser (vers. 2)	184d						
T2.1 - Review and identification of appropriate technologies for semantic interpretation	62d						
T2.2a - Definition of a common domain for the legislative corpus	62d						
T2.2b - Production of a multilingual annotated corpus within the common domain	181d						
T2.3a - Development of a prototype for the automatic classification of provisions (vers. 1)	122d						
T2.3b - Improvement of prototype for the automatic classification of provisions (vers. 2)	184d						
T2.4a - Development of language-specific prototypes for the provision roles annotation (v. 1)	184d						
T2.4b - Improvement of language-specific prototypes for the provision roles annotation (v. 2)	122d						
T3 - Development of a prototype of distributed resolution system	305d						
<b>WP 4 - Conceptual indexing of normative documents</b>	<b>793d</b>						
T1 - Import of multilingual lexicons and core ontologies for Law	184d						
T2 - Term extraction from significant elements of provisions	184d						
T3 - Localization of the ontology in the languages of the consortium & multilingual mapping	181d						
T4 - Mapping of terms onto the ontology and document indexing	213d						
T5 - Tools for bottom-up, semi-automatic building of multilingual domain-ontologies	366d						
<b>WP 5 - Integrated workbench for the editing</b>	<b>915d</b>						
T1 - Review and identification of appropriate technologies	62d						
T2a - Prototype of an integrated workbench for the editing of new texts (vers. 1)	395d						
T2b - Prototype of an integrated workbench for the editing of new texts (vers. 2)	488d						
T3a - Prototype of an integrated workbench for the editing of legacy texts (vers. 1)	365d						
T3b - Prototype of an integrated workbench for the editing of legacy texts (vers. 2)	488d						
<b>WP 6 - Methods and tools for the quality control</b>	<b>915d</b>						
T1 - Survey of evaluation parameters (25 EU countries)	62d						
T2 - Definition of a set of common parameters for normative texts quality evaluation	150d						
T3 - Implementation or conversion of factual data base	215d						
T4 - Development of a prototype for the automatic evaluation	274d						
T5 - Translation of the evaluation into statistical indexes	184d						
<b>WP 7 - Multilingual semantic search and retrieval</b>	<b>731d</b>						
T1a - Definition of system functionality, user interface, ranking & evaluation criteria	120d						
T1b - Integrate linguistic processor & domain wordnets for the test data languages	182d						
T2a - Collection of the test data with the enriched metadata	91d						
T2b - Build the conceptual indexes and the meta indexes	92d						
T2c - Retrieval for combined cross-lingual conceptual & metadata search	92d						
T3a - Extraction and translation of the test queries	92d						
T3b - Extraction of the end-user scenarios	92d						
<b>WP 8 - Automatic reconstruction of the normative text in force</b>	<b>488d</b>						
T1a - Construction of meaningful test data set for whole partition amendments	122d						
T1b - Prototype of text in force reconstruction for textual amendments of whole partitions	182d						
T2a - Construction of meaningful test data set for partition subset amendments	182d						
T2b - Prototype of text in force reconstruction for textual amendments of partition subsets	184d						
<b>WP 9 - Evaluation</b>	<b>1096d</b>						
T1 - Specification of evaluation methodology & criteria, test & development corpora	181d						
T2a - Evaluation of WP2 results (vers. 1)	92d						
T2b - Evaluation of WP2 results (vers. 2)	91d						
T3a - Evaluation of WP3 results (vers. 1)	92d						
T3b - Evaluation of WP3 results (vers. 2)	91d						
T4 - Evaluation of WP4 results	91d						
T5 - Definition of evaluation criteria in demonstrative applications	182d						
T6 - Evaluation of WP7 results	92d						
T7 - Evaluation of WP8 results	184d						
T8 - Continuous collection of feedback and refinement of evaluation criteria	1096d						
<b>WP 10 - Information dissemination and exploitation</b>	<b>1096d</b>						
T1 - Actions by each "public partner" to gain involvement by the interested PAs	731d						
T2 - Organization of a pan European conference about language resources and tools	122d						
T3 - Participation to national events and European events	731d						
T4 - Creation of a web site with demonstrators and activation of personal contacts	1096d						
T5 - Production of a detailed exploitation plan, towards PAs & publishers of legal texts	488d						
T6 - Publication of papers on specialized magazines	731d						
T7 - Presentation of papers to specialized conferences on milestones & expected result	731d						
T8 - Creation of a user forum	731d						

b) Graphical presentation of the components, showing their interdependencies



d) Detailed work description broken down into workpackages:

### Workpackage list

Work-package No <sup>1</sup>	Workpackage title	Lead contractor No <sup>2</sup>	Person-months <sup>3</sup>	Start month <sup>4</sup>	End month <sup>5</sup>	Deliverable No <sup>6</sup>
WP1	Management and coordination	1	26	0	36	D1.1-D1.7
WP2	Standardized representation of normative documents	2	36,5	0	24	D2.1-D2.6
WP3	Automatic normative document annotation	4	87	6	24	D3.1-D3.11
WP4	Conceptual indexing of normative documents	1	80	6	27	D4.1-D4.3
WP5	Integrated workbench for the editing of normative texts	5	58	6	36	D5.1-D5.2
WP6	Methods and tools for the quality control of normative texts	1	31,5	6	36	D6.1-D6.2
WP7	Multilingual semantic search and retrieval of normative documents	9	27,5	12	36	D7.1-D7.2
WP8	Automatic reconstruction of the normative text in force	6	40,5	20	36	D8.1-D8.2
WP9	Evaluation	7	21,5	1	36	D9.1-D9.3
WP10	Information dissemination and exploitation	13	50,5	1	36	D10.1-D10.3
	TOTAL		459			

<sup>1</sup> Workpackage number: WP 1 – WP n.

<sup>2</sup> Number of the contractor leading the work in this workpackage.

<sup>3</sup> The total number of person-months allocated to each workpackage.

<sup>4</sup> Relative start date for the work in the specific workpackages, month 0 marking the start of the project, and all other start dates being relative to this start date.

<sup>5</sup> Relative end date, month 0 marking the start of the project, and all ends dates being relative to this start date.

<sup>6</sup> Deliverable number: Number for the deliverable(s)/result(s) mentioned in the workpackage: D1 - Dn.

(Deliverables list, use Deliverables list form below)

### Deliverables list

Deliverable No <sup>7</sup>	Deliverable title	Delivery date <sup>8</sup>	Nature <sup>9</sup>	Dissemination level <sup>10</sup>
D1.1	Project management plan	2	R	PU
D1.2	1 <sup>st</sup> Periodic Progress Report	6	R	PU
D1.3	2 <sup>nd</sup> Periodic Progress Report	12	R	PU
D1.4	3rd Periodic Progress Report	18	R	PU
D1.5	4th Periodic Progress Report	24	R	PU
D1.6	5th Periodic Progress Report	30	R	PU
D1.7	Final Report	36	R	PU
D2.1	Core standard for identification and structural description of normative documents	6	R	PU
D2.2	Extended standard for identification and structural description of normative documents	12	R	PU
D2.3	Core standard for semantic description of normative documents	6	R	PU
D2.4	Extended standard for semantic description of normative documents	12	R	PU
D2.5	Standard for identification and structural description of normative documents	24	R	PU

<sup>7</sup> Deliverable numbers in order of delivery dates: D1 – Dn

<sup>8</sup> Month in which the deliverables will be available. Month 0 marking the start of the project, and all delivery dates being relative to this start date.

<sup>9</sup> Please indicate the nature of the deliverable using one of the following codes:

**R** = Report

**P** = Prototype

**D** = Demonstrator

**O** = Other

<sup>10</sup> Please indicate the dissemination level using one of the following codes:

**PU** = Public

**PP** = Restricted to other programme participants (including the Commission Services).

**RE** = Restricted to a group specified by the consortium (including the Commission Services).

**CO** = Confidential, only for members of the consortium (including the Commission Services).

D2.6	Standard for semantic description of normative documents	24	R	PU
D2.3	Specifications for the semantic description of normative documents	10	R	PU
D3.1.	Language-independent prototype of a cross-reference parser engine	12	P	RE
D3.2	Language-independent prototype of structural parsing engine	12	P	RE
D3.3	Language-dependent prototypes for cross-reference parsing	24	P	RE
D3.4	Language-dependent prototypes for structural parsing	24	P	RE
D3.5	Prototype of the automatic classifier of normative documents	24	P	RE
D3.6	Prototype of the automatic semantic roles extractor for Italian	24	P	RE
D3.7	Prototype of the automatic semantic roles extractors for Dutch	24	P	RE
D3.8	Prototype of the automatic semantic roles extractor for Hungarian	24	P	RE
D3.9	Prototype of the automatic semantic roles extractor for Danish	24	P	RE
D3.10	Prototype of the automatic semantic roles extractor for Greek	24	P	RE
D3.11	Prototype of the distributed resolution system	18	P	RE
D4.1	Report on multilingual domain ontology	16	R	PU
D4.2	Report on automatic term extraction	20	R	PU
D4.3	Report on normative document conceptual indexing	27	R	PU
D5.1	First version of the integrated workbench for the editing of normative texts	20	P	RE
D5.2	Final version of the integrated workbench for the editing of normative texts	36	P	PU

D6.1	Specifications of a standard for the quality evaluation of the normative texts	19	R	PU
D6.2	Application prototype.	30	P	RE
D7.1.	Application prototype	32	D	RE
D7.2	Evaluation report	36	R	PU
D8.1.	Prototype application for the reconstruction of the text-in-force (textual amendments of whole partitions)	30	D	RE
D8.2	Prototype application for the reconstruction of the text-in-force (textual amendments of partitions subsets)	36	D	RE
D9.1	Specification of evaluation methodology, evaluation criteria, test corpora and development corpora	6	R	PU
D9.2	First evaluation report	18	R	PU
D9.3	Final evaluation report	36	R	PU
D10.1	Project Advisory Panel	12	O	RE
D10.2	Maintenance of web site and email lists. Lists of actors and relevant organisations. Dissemination and use plan. Project presentation	2	O	RE, PU
D10.3	Dissemination and exploitation plan	9	R	PU
D10.4	Dissemination and awareness campaign report.	36	R	PU



**Workpackage description**

Workpackage number	WP1	Start date or starting event:											Month 0
Workpackage title <i>Management and coordination</i>													
Participant id	1	2	3	4	5	6	7	8	9	10	11	12	13
Person-months per participant:	20	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

**Objectives**

- Coordinate relationships between partners of the Consortium, ensuring timely transfer of specifications, tools etc. between WPs;
- Monitor and track progress advancement in the project (schedules, milestones, deliverables), resource allocation and spending;
- Swift resolution of unexpected problems, ensure Project plan refinements if needed;
- Handle administration and financial management of the project;
- Organise and lead project meetings.

**Description of work**

CNR will perform the following organisational tasks:

Personnel of CNR-ILC taking care of scientific and organisational issues will take care of the following:

- Scheduling of activities related to the Contract, and coordination with key personnel of each partner in the Consortium;
- Scheduling of activities according to the project plan and in agreement with key personnel of each partner in the Consortium;
- Gathering of deliverables from partners and assisting them in the writing of quarterly reports about the progress status of the project;
- Organisation of quarterly meetings between partners to share solutions, critical issues, and results;
- Organisation of working meetings/extraordinary meetings between all partners or specific partners, when necessary;
- Organising public events (press conferences, workshops, public meetings);
- Organising and performing the Project Presentation.

Personnel of CNR-ITTIG will take care of the following financial and administration activities:

- Compile periodic certifications as requested by the EC;
- Organise working meetings among administration personnel (also using videoconferencing systems) each semester to share problems, issues and related

solutions;

- Organise extraordinary meetings among all or part of administrative staff (also using videoconferencing systems), when necessary;
- Organise and manage audits;
- Administrative and financial reporting of the project;
- Manage relations and necessary contacts with the EC.

See also section on project management in Part B5.

### **Deliverables**

D1.1 Project management plan

D1.2 1<sup>st</sup> Periodic Progress Report

D1.3 2nd Periodic Progress Report

D1.4 3rd Periodic Progress Report

D1.5 4th Periodic Progress Report

D1.6 5th Periodic Progress Report

D1.7 Final Report

### **Milestones<sup>11</sup> and expected result**

This WP is a horizontal activity across all other WPs, and therefore connected to each one.

Activities of this WP do not impact directly on the specific implementation or realization of tasks of the other WPs, but are designed to support them.

The following milestones are foreseen in addition to Project Reports: project review and project management plan (M2); review of project development, assessment of deviations from project plan and design of recovery measures, if necessary (M12+1) and (M24+1); final financial report of the project (M36+2).

Expected results: times and schedules met; reviews by the committee with positive results; quarterly certifications and final administrative reporting of the project with positive results.

---

<sup>11</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	WP2		Start date or starting event:									Month 1			
Workpackage title <i>Standardized representation of normative documents</i>															
Participant id	2	1	3	4	6	7	10	11	12	13					
Person-months per participant:	9	10	1,5	4	3,5	2	2	1,5	2	1					

## Objectives

To develop a standard representation formalism for the identification of normative documents (both at the European and national level) as well as the description of their formal and semantic structure.

## Description of work

T1: *Review of state of the art* in representation formalisms for the unique identification of normative documents, the description of their formal structure and the representation of their semantic content. Particular attention will be given to existing initiatives inside the EU, with the aim of a thorough revision of relevant initiatives across all 25 EU countries.

T2: Specification of a standard schema document (using DTDs, XML Schema, Relax NG) for an XML-based formalism for a) the unique identification and b) the structural description of normative documents. The purpose of the first standard is to provide a means for univocally identifying any legal measure issued by any European authority as well as providing a way to express references in a stable way independently of document physical locations.

The standard for structural description of documents aims at endowing documents with a representation of their formal structure able to describe all the meaningful partitions in legal measures of different countries. This task is further divided in two subtasks to:

T2.1 Identify the common structural aspects shared by normative documents of all EU countries (hierarchies, titles, references, annexes, etc.) and express them in a *core schema* usable in all different countries.

T2.2 Provide some exemplifying instantiations of the core schema to account for the structural aspects that are individual of each specific document type, both across and within national and language borders, and express them in a number of *additional schemas* to be combined with the core schema. The schemas will be specially targeted at least for two of the languages and nationalities available within the Consortium.

T3: Specification of a standard for the semantic description of normative documents. Semantic description of normative documents will be achieved by means of two types of metadata or descriptors: *general* and *analytic* metadata, to be specified as an RDF Schema/OWL ontology. This task is further divided in two subtasks to:

T3.1 describe general facts about the documents (general metadata), such as their publication, their relation to other norms, their role and status, their history and lifecycle.

T3.2 describe analytically the semantic content of the normative documents, i.e. the particular provisions expressed therein. To this end, we expect to adopt a representation model organised into types of provisions (i.e. obligation, permission, prohibition, etc.). Moreover, each type of provision will be further modelled as to express the particular semantic roles relevant for each type (such as, for instance, the addressee and obliged action of an obligation, the sanctioned action and the foreseen penalty in a sanction, etc.). In

designing a model for analytical metadata particular attention will be given to usability for search purposes.

Again, the standard in T3 will be developed around a *core schema* organised into macro-classes of provisions and covering a minimum common semantic framework shared among the different legislative systems. In addition to the core schema, a set of exemplifying instantiations will be provided (extended schema), where the macro-classes are further articulated into more refined classes or types taking into account local legislative practices.

The two standards will be iteratively refined on the basis of user feedback and an improved version will be issued at month 24.

### **Deliverables**

D2.1 Core standard for identification and structural description of normative documents (due M6).

D2.2 Extended standard for identification and structural description of normative documents (due M12).

D2.3 Core standard for semantic description of normative documents (due M6).

D2.4 Extended standard for semantic description of normative documents (due M12).

D2.5 Standard for identification and structural description of normative documents (due M24)

D2.6 Standard for semantic description of normative documents (due M24).

### **Milestones<sup>12</sup> and expected result**

M1 (month 6): delivery of the core version of the URN schema

M2 (month 6): delivery of the core XML schema module for document structure

M3 (month 6): delivery of the core RDF/OWL module for general and analytical semantic metadata

M4 (month 12): delivery of extended version of the URN schema

M5 (month 12): delivery of the extended XML schema module for document structure

M6 (month 12): delivery of the extended RDF/OWL module for general and analytical semantic metadata

M7 (month 24): delivery of the URN schema (improved version)

M8 (month 24): delivery of the XML schema module for document structure (improved version)

M9 (month 24): delivery of the RDF/OWL module for general and analytical semantic metadata (improved version)

<sup>12</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	WP3	Start date or starting event:									Month 6			
Workpackage title <i>Automatic normative document annotation</i>														
Participant id	4	1	2	5	6	7	8	9	10	11				
Person-months per participant:	17	20	5,5	7	2,5	9	12	11	1,5	1,5				

### Objectives

To develop instruments for the automatic qualification of normative documents according to the identification, formal and semantic standards developed in WP2.

### Description of work

The work in WP3 is organised around three main tasks:

T1 Automatic legacy content handling is devoted to the development of tools able to detect the formal profile of a normative document and produce its XML description compliant with the standard schema formalism developed in WP2. The following sub-tasks can be envisaged:

T1.1: Review and identification of appropriate technologies

T1.2: Definition of a common set of normative documents to act as a development corpus

T1.3a: Development of a language-independent prototype of a cross-reference parser engine

T1.3b: Development of a language-independent prototype of structural parsing engine

T1.4a: Development of language-dependent prototypes for cross-reference parsing (first version)

T1.4b: Improvement of language-dependent prototypes for cross-reference parsing (second version)

T1.4c: Development of language-dependent prototypes for structural parsing (first version)

T1.4d: Improvement of language-dependent prototypes for structural parsing (second version)

T2 Automatic semantic annotation refers to the development of algorithms and tools for the automatic semantic interpretation of normative documents, i.e. the assignment of portions of normative documents to provision types and extraction of text strings corresponding to relevant semantic roles. The following sub-tasks can be envisaged:

T2.1: Review and identification of appropriate technologies for semantic annotation

T2.2: Definition of a common domain and production of a multilingual annotated corpus

T2.3a: Development of a prototype for the automatic classification of law paragraphs according to the provision classes as specified by WP2 (first version)

T2.3b: Improvement of a prototype for the automatic classification of law paragraphs (second version)

T2.4: Development of language-specific prototypes for the automatic annotation of normative documents with provision-specific semantic roles. Prototypes will be implemented for Italian, Dutch, Hungarian, Danish and Greek.

T3 Development of a prototype of a distributed resolution system

**Deliverables**

- D3.1: Language-independent prototype of a cross-reference parser engine (due M12)
- D3.2: Language-independent prototype of structural parsing engine (due M12)
- D3.3: Language-dependent prototypes for cross-reference parsing (due M24)
- D3.4: Language-dependent prototypes for structural parsing (due M24)
- D3.5: Prototype of the automatic classifier of normative documents (due M24)
- D3.6: Prototype of the automatic semantic roles extractor for Italian (due M24)
- D3.7: Prototype of the automatic semantic roles extractors for Dutch (due M24)
- D3.8: Prototype of the automatic semantic roles extractor for Hungarian (due M24)
- D3.9: Prototype of the automatic semantic roles extractor for Danish (due M24)
- D3.10: Prototype of the automatic semantic roles extractor for Greek (due M24)
- D3.11: Prototype of the distributed resolution system (due M18)

**Milestones<sup>13</sup> and expected result**

- M1 (month 8): the corpus of normative documents has been identified and ready to be used for development of the cross-reference and structural parsers.
- M2 (month 13): language-independent prototypes of cross-reference and structural parsing engine ready.
- M3 (month 14): production of the multilingual annotated corpus for the development of the semantic annotation modules.
- M4 (month 18): first version of language-dependent prototypes for cross-reference and structural parsing ready.
- M5 (month 18): first version of prototype for automatic classification of provisions ready.
- M6 (month 18): prototype of distributed resolution system ready.
- M7 (month 20): first version of prototypes for provision roles annotation ready for Italian, Dutch, Hungarian, Danish and Greek.
- M8 (month 24): second version of language-dependent prototypes for cross-reference and structural parsing ready.
- M9 (month 24): second version of prototype for automatic classification of provisions ready.
- M10 (month 24): second version of prototypes for provision roles annotation ready for Italian, Dutch, Hungarian, Danish and Greek.

---

<sup>13</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	WP4	Start date or starting event:										Month 6			
Workpackage title <i>Conceptual indexing of normative documents</i>															
Participant id	1	2	4	6	7	8	9	10	11	12					
Person-months per participant:	20	5,5	14	4,5	10	8	11	1,5	1,5	4					

### Objectives

The principal aim of this WP is the development of algorithms and tools for the automatic and semi-automatic processing of multilingual normative data for the twofold end of:

- c) using normative documents as a *term repository* which can be exploited to populate domain-specific ontologies with data belonging to real texts, thus reducing the costly process of manual ontology building;
- d) *producing a conceptual indexing of documents* with semantic metadata corresponding to the terms, concepts and relations of the ontologies.

To this end WP4 will exploit semantic web techniques for the automatic and semi-automatic identification of the conceptual entities addresses/evoked by the normative documents; it will identify and import appropriate existing multilingual ontologies, most probably based on the WordNet model. The adopted ontology will be expanded for the domain chosen and localized in the necessary languages.

### Description of work

T1: Import of multilingual lexicons and core ontologies for Law (e.g., JurWordNet): the project will privilege the re-use and integration of the proposals in the field of multilingual lexicons for law and of legal ontologies, for instance multilingual localisation of legal WordNets.

T3: Localization of the ontology in Italian, Greek, Danish, Hungarian and Dutch and construction of the multilingual mapping

T2: Automatic Term extraction from normative documents. Potential terms and their variants will be automatically extracted from a corpus of normative documents to be set out for specific use in this WP and covering Italian, Greek, Danish, Hungarian and Dutch. The algorithms for automatic term extraction will combine language-independent and language-dependent methods so as to ensure commonality across the tools developed for the various languages while allowing for integration of language-specific NLP tools already developed by the involved partners.

T4: Mapping of terms onto the ontology and document indexing: the terms extracted automatically will then be semi-automatically mapped onto the multilingual ontology developed under T1-T2. This conceptual indexing will be added up to the type of semantic indexing developed under WP3, so as to provide normative documents with an in-depth representation of the normative knowledge expressed.

T5: Development of tools for bottom-up, semi-automatic building of multilingual domain-ontologies. WP4 will also develop algorithms and tools helping in the bottom-up, semi-automatic building domain-ontologies on the basis of the terms extracted from normative documents and the relations discovered among terms. These tools will also help in T4.

**Deliverables**

D4.1 Report on multilingual domain ontology (due M 16)

D4.2 Report on automatic term extraction (due M 20)

D4.3 Report on normative document conceptual indexing (due M 27)

**Milestones<sup>14</sup> and expected result**

M1 (month 18): domain-specific domain ontology fully localized.

M2 (month 12): corpus of domain-specific normative documents ready.

M3 (month 19): automatic extraction of terms from corpus complete.

M4 (month 24): extracted terms mapped onto the ontology and ontologies extended with new terms.

M5 (month 27): indexing of the corpus of normative documents complete.

---

<sup>14</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.



Workpackage number	WP5	Start date or starting event:								Month 1	
Workpackage title <i>Integrated workbench for the editing of normative texts</i>											
Participant id	5	1	2	3	4	6	7	10	11	12	13
Person-months per participant:	20	6	3	0,5	12	1	2	0,5	1,5	2	19,5

### Objectives

To realize a workbench integrating the various modules/functionalities for the automatic annotation of normative documents, supplemented by functionalities for the editing/drafting of new texts.

### Description of work

In order to make easier the adoption of the established European legal standards, an editor operating within the URN and XML-DTDs or XML-Schema framework established by the project will be developed.

T1: Review and identification of the appropriate technologies

T2a: Development of a prototype for integrated workbench for the editing of new normative documents (first version);

T2b: Improvement of a prototype for integrated workbench for the editing of new normative documents (second version);

T3a: Development of a prototype for integrated workbench for the editing of legacy normative documents (first version);

T3b: Improvement of a prototype for integrated workbench for the editing of legacy normative documents (second version);

### Deliverables

D5.1 First version of the integrated workbench for the editing of normative texts (due month 20)

D5.2 Final version of the integrated workbench for the editing of normative texts (due month 36)

### Milestones<sup>15</sup> and expected result

M1 (month 18): first version of prototype of an integrated workbench for the editing of new normative documents ready for evaluation;

M2 (month 18): first version of prototype of an integrated workbench for the editing of legacy normative documents ready for evaluation;

M3 (month 33): second version of prototype of an integrated workbench for the editing of new normative documents ready for evaluation under WP9

M4 (month 33): second version of prototype of an integrated workbench for the editing of legacy

<sup>15</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

normative documents ready for evaluation under WP9
--

Workpackage number	WP6	Start date or starting event:								Month 6			
Workpackage title <i>Methods and tools for the quality control of normative texts</i>													
Participant id	1	3	4	5	6	7	10	11	12				
Person-months per participant:	12	0,5	2	10	1,5	2	0,5	1	2				

**Objectives**

To provide instruments and methodologies for evaluating the quality of normative texts.

**Description of work**

T1: Survey of evaluation parameters across 25 countries;

T2: Definition of a set of common parameters for the quality evaluation of the normative texts, for the EU countries (*law quality indicators*);

T3: Implementation or conversion of factual data base;

T4: Development of a prototype for the automatic evaluation;

T5: Translation of the evaluation into statistical indexes (*law quality indices*).

**Deliverables**

D6.1 Specifications of a standard for the quality evaluation of the normative texts (due M19)

D6.2 Application prototype (due M30)

**Milestones<sup>16</sup> and expected result**

M1 (month 14): parameters for normative text quality evaluation identified

M2 (month 20): conversion of factual data base.

M3 (month 30): prototype for automatic evaluation of normative texts ready.

M4 (month 36): translation of the evaluation into statistical indexes

<sup>16</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	WP7	Start date or starting event:								Month 12			
Workpackage title <i>Multilingual semantic search and retrieval of normative documents</i>													
Participant id	9	1	3	4	7	10	11	12	13				
Person-months per participant:	11	5	0,5	2	2	0,5	1,5	2	3				

### Objectives

The objective of this WP is to demonstrate the usefulness and usability of semantically annotated normative data through the development of an advanced search and retrieval application. The retrieval system will combine meta search with conceptual indexes on free text. The meta-search will be applied to the automatically extracted meta-information from the legal documents. The conceptual indexes will be built with wordnets that are tuned to the legal domain. The search will be cross-lingual covering all the project languages. Finally, the retrieval system will show activation of inferences for the retrieval of implicit and/or related norms (logical and legislative inferences). The demonstration will measure the productivity of the added value for the legal domain. It will include:

- a precision/recall benchmark with a substantial number of queries derived from the corpus and translated to all the project languages
- a end-user evaluation by a number of end-users, based on pre-designed retrieval sceneries. The end-user behaviour is logged and used for measuring effectiveness.

The demonstration will use the TwentyOne retrieval system developed by Irion Technologies.

### Description of work

#### T1. Preparatory work:

T1a: Design of the evaluation criteria for retrieval

T1b: Integrate the linguistic processors and the domain wordnets for the legal domain for the languages of the test data in the TwentyOne System

#### T2: Building the demonstrator:

T2a: Collection of the test data with the enriched meta data

T2b: Build the conceptual indexes and the meta indexes

T3b: Retrieval interface for combined cross-lingual conceptual search and meta-search

#### T3: Demonstration:

T3a: Extraction and translation of the test queries

T3b: Extraction of the end-user sceneries

T4: Evaluation of the test queries and the end-user tasks

### Deliverables

D7.1 Application prototype (due M 32)

D7.2 Evaluation report (due M36)
----------------------------------

<b>Milestones<sup>17</sup> and expected result</b>
--

M1 (month 16): definition of system functionality, user interface, ranking and evaluation criteria complete
---

M2 (month 32): demonstrator fully functional.
---

---

<sup>17</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	WP8	Start date or starting event:										Month 20					
Workpackage title Automatic reconstruction of the normative text in force																	
Participant id	6	1	2	3	4	5	7	10	11	12							
Person-months per participant:	4	5	3	0,5	2	20	2	0,5	1,5	2							

**Objectives**

Demonstrate the usefulness and usability of structurally annotated normative data through the development of an application for the automatic reconstruction of the normative text in force.

**Description of work**

The concept will be proved through realisation of an application for the automatic reconstruction of the normative text through its modification history, when modifications apply to entire partitions. The application will be developed for all languages of the Consortium.

T1: Textual amendments for whole partitions:

- a) Construction of a meaningful test data set;
- b) Development of a prototype of reconstruction of the text in force;

T2: Textual amendments for subsets (words, phrases, etc.) of partition:

- a) Construction of a meaningful test data set;
- b) Development of a prototype of reconstruction of the text in force;

**Deliverables**

D8.1: Prototype application for the reconstruction of the text-in-force (textual amendments of whole partitions) (due M30)

D8.2: Prototype application for the reconstruction of the text-in-force (textual amendments of partitions subsets) (due M36)

**Milestones<sup>18</sup> and expected result**

M1 (month 29): prototype the reconstruction of the text-in-force limited to textual amendments of whole partitions ready for evaluation

M2 (month 35): prototype the reconstruction of the text-in-force extended to textual amendments of partitions subsets ready for evaluation

<sup>18</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

Workpackage number	WP9	Start date or starting event:								Month 1				
Workpackage title Evaluation														
Participant id	7	1	3	4	6	10	11	12	13					
Person-months per participant:	6	2	1,5	1	1	3	3	2	2					

### Objectives

Perform extensive functionality and usability testing of standards for structural and semantic annotation of normative data, both independently and embedded into the demonstrative applications.

### Description of work

WP9 will set up evaluation procedures early in the project, especially by gathering input from users' requirements. The work will build on existing standards for quality and quality assurance, e.g. the ISO 9000 series. It will continuously evaluate WP2, WP3 and WP4 progress and results. WP9 will take into account everything of relevance from the feedback received during the WP10 dissemination campaign, including early feedback on WP2, then on WP2 specifications, then on the first application prototype, and WP9 will design and do the evaluation of the demonstrative applications to be developed in WP7 and WP8.

The *Advisory Group* created in WP10 will give input and advice on formalisms, annotation schemes, platforms, functionalities and usability.

A central role within the *Advisory Group* will be played by the *Official Journal of the European Union* (OJ).

Moreover *Advisory Group* will comprise industrial users (SMEs), professional associations, dedicated publishers, PA representatives.

Taking into account the input of the *Advisory Group* and the WP2 specification a preliminary list of evaluation criteria will be defined. In the early phase of the project a test corpus of legal texts will be collected, together with a development corpus to be used in WP3 and WP4.

### Tasks

T1 Specification of evaluation methodology, evaluation criteria, test corpora and development corpora.

T2 Evaluation of WP2 results.

T3 Evaluation of WP3 results

T4 Evaluation of WP4 results

T5 Definition of evaluation criteria in demonstrative applications

T6 Evaluation of WP7 results

T7 Evaluation of WP8 results

T8 Continuous collection of feedback and refinement of evaluation criteria

### Deliverables

D9.1 Specification of evaluation methodology, evaluation criteria, test corpora and development corpora (due M 6)

D9.2 First evaluation report (due M 18)

D9.3 Final evaluation report (due M36)

### **Milestones<sup>19</sup> and expected result**

The milestones of this WP are naturally appearing at the two evaluation reports.

The expected result of having an integrated and yet independent quality assurance and evaluation activity going on throughout the project is that user requirements will be taken into account at a high level, and that progress can be measured continuously and according to agreed standards, thereby ensuring a high quality level of the project deliverables.

---

<sup>19</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.



Workpackage number	WP10	Start date or starting event:										Month 1		
Workpackage title Information dissemination and exploitation														
Participant id	13	1	2	3	4	5	6	7	8	9	10	11	12	
Person-months per participant:	19,5	3	2	4	1	1	2	1	1	1	5	5	5	

### Objectives

The objectives of this WP are twofold:

For dissemination, to create and maintain the project website, produce public releases of the project Editor and recommendations, build a User Forum, publicise results widely.

For exploitation, the goal of this WP is to put in place all possible measures in order to create the necessary consensus for early exploitation of EUNOMIA resources. At the end of this WP, the consortium should have created a group of interested parties that are ready to bring forward all the necessary activities on a European scale and to develop cost-effective, customer-oriented applications. Such a group will include public administration, service provider for public administration, private publishers and entities able to carry on the task of full localization of the editor to languages that are not covered by EUNOMIA.

### Description of work

In order to achieve the described objectives we plan the following activities:

T1: Create a project Advisory Group which will be set out at the beginning of the project and will be gathering industrial users (SMEs), professional associations, dedicated publishers, PA representatives, with the purpose of giving input and advice on formalisms, annotation schemes, platforms, functionalities and usability.

T2. Actions by each “public partner” in order to gain involvement by the public administrations they are interfacing with.

T3. Organization of a pan European workshop/conference about language resources and tools for easy information access in the legal domain.

T4. Participation to national events and European Events such as the Italian *Forum P.A* (<http://www.forumpa.it>) or the *EU Quality Conference for Public Administrations*.

T5. Creation of a web site with demonstrators and activation of personal contacts with publishing houses in order to create an active e-community on the subject of language enabled access to legal documentation.

T6. Production of a detailed exploitation plan. Such an exploitation plan will be divided into two sections, i.e. the one oriented towards public administrations and the one oriented towards publishers of legal texts.

T7. Publication of papers on magazines specialized for publishers, for legal informatics, and for the public sector.

T8. Presentation of papers to specialized conferences such as the *JURIX Conference*.  
(inter-) Dependencies, milestones and expected result

T9. Creation of a User Forum, involving citizens, PAs, publishers and service providers.

### Deliverables

D10.1 Project Advisory Group (due M12).

D10.2 Maintenance of web site and email lists. Lists of actors and relevant organisations. Project

presentation. (due M2 onwards).  
D10.3 Dissemination and exploitation plan (due M9)  
D10.4 Dissemination and awareness campaign report. (due M 36)

### **Milestones<sup>20</sup> and expected result**

M1 (month 3): Project website is operational.  
M2 (month 9): Dissemination and exploitation plan.  
M3 (month 12): Advisory Panel is fully functional.  
M4 (month 36): Dissemination and awareness campaign report.

## **B.7 Other issues**

No recommended length – depends on the number of such other issues which the project involves

### ***B.7.1 Ethical and gender issues***

#### **Ethical issues checklist**

**Table A. Proposers are requested to fill in the following table**

<b>Does your proposed research raise sensitive ethical questions related to:</b>	<b>YES</b>	<b>NO</b>
Human beings		X
Human biological samples		X
Personal data (whether identified by name or not)		X
Genetic information		X
Animals		X

*If you answer “YES” to any of the above, please include in your proposal section B7.1 the more detailed version of Table A (“Crucial information”) obtained from: [http://europa.eu.int/comm/research/science-society/ethics/rules\\_en.html](http://europa.eu.int/comm/research/science-society/ethics/rules_en.html) and also incorporate in section B.7.1 and in other appropriate parts of your proposal comments corresponding to the detailed instructions given in sections C-D at the above address*

<sup>20</sup> Milestones are control points at which decisions are needed; for example concerning which of several technologies will be adopted as the basis for the next phase of the project.

**Table B. Proposers are requested to confirm that the proposed research does not involve:**

Research activity aimed at human cloning for reproductive purposes,  
 Research activity intended to modify the genetic heritage of human beings which could make such changes heritable<sup>21</sup>  
 Research activity intended to create human embryos solely for the purpose of research or for the purpose of stem cell procurement, including by means of somatic cell nuclear transfer.

<b>Confirmation : the proposed research involves none of the issues listed in Table B</b>	<b>YES</b>	<b>NO</b>
	<b>X</b>	

*Further information on ethics requirements and rules are given at the science and ethics website at [http://europa.eu.int/comm/research/science-society/ethics/ethics\\_en.html](http://europa.eu.int/comm/research/science-society/ethics/ethics_en.html)*

---

<sup>21</sup> Research relating to cancer treatment of the gonads can be financed

***B.7.2 Other EC-policy related issues***

No recommended length – depends on the number of such other issues which the project involves

*(Are there other EC-policy related issues, and are they taken into account? Demonstrate a readiness to engage with actors beyond the research to help spread awareness and knowledge and to explore the wider societal implications of the proposed work; if relevant set out synergies with education at all levels.)*