

Global Centre for ICT in Parliament Working Paper No. 2

January 2008

Legal Informatics and Management of Legislative Documents

Authors:

Mariangela Biasiotti (ITTIG, Florence)

Enrico Francesconi (ITTIG, Florence)

Monica Palmirani (CIRSFID, University of Bologna)

Giovanni Sartor (EUI, Florence)

Fabio Vitali (Department of Computing, University of Bologna)

Editor: Giovanni Sartor

Abstract

This report aims at providing an overview of the state of the art and of the prospects of the application of Information and Communication Technologies (ICT) in the legislative domain, in particular concerning the management of legislative documents. After a brief introduction on legal informatics, we focus on legislative informatics and identify the challenges it faces nowadays, in the framework of the Internet and the (semantic) Web. We then describe the evolution of the ICT-based management of legislative documents and identify and evaluate the emergent approaches, focusing on those based on open standards. The report is completed by two appendixes: the first reviews initiatives pertaining to the standard-based management of legal sources, the second reviews initiatives pertaining to semantic resources for legislation.

Disclaimer

The Working Papers of the Global Centre for ICT in Parliament are preliminary documents circulated in a limited number of copies and posted on the Global Centre website at <http://www.ictparliament.org> to stimulate discussion and critical comment. The views and opinions expressed herein are those of the author and do not necessarily reflect those of the Global Centre for ICT in Parliament. The designations and terminology employed may not conform to United Nations practice and do not imply the expression of any opinion whatsoever on the part of the Organization.

Table of Contents

1. Introduction: ICT and legislation in the knowledge society	4
1.1 ICT and the predicaments of legislation	4
1.1.1 Predicaments of legislation	4
1.1.2 The role of ICT	5
1.1.3 Legislative information systems	8
1.2 Legal and legislative informatics	10
1.2.1 A short history of legal informatics	10
1.2.2 From legal to legislative informatics	12
2 Legislative information and the Web (Giovanni Sartor)	14
2.1 Trends in the provision of legal information	14
2.1.1 The digitalization of legal information	14
2.1.2 The law on the Internet	15
2.1.3 The standardization of formats for legal information	15
2.1.4 The law in the semantic Web	16
2.1.5 The executable representation of legal information	16
2.1.6 An increasing diversity in information providers	17
2.2 Focus on the semantic Web	17
2.2.1 The semantic Web	17
2.2.2 The legal semantic Web	20
3 ICT-based management of legislative documents (Giovanni Sartor)	23
3.1 A short history of legal information systems	23
3.1.1 The beginnings	24
3.1.2 The thesis: The remotely accessible centralized systems of the 60's and the 70's	24
3.1.3 The antithesis: The multiple isolated systems of the 1980s	25
3.1.4 The synthesis: The universally accessible but plural systems of the 1990s	27
3.1.5 Legal information institutes	29
3.2 The dialectics of current approaches	30
3.2.1 A new thesis: A centralized legal information system for the Internet age	31
3.2.2 A new antithesis: Access to distributed legal resources	33
3.2.3 A new synthesis: Standard-based legal information in the jurisphere	35
3.2.4 Standard-based management of legislative documents: The Norme-in-rete example	37
3.2.5 Standard-based management of legislative documents: The AKOMA-NTOSO example	39
4 A standard-based approach for the management of legislative documents (Fabio Vitali)	42
4.1 The objective of a standard for legal documents	42
4.1.1 A rationale for shared standards	43
4.1.2 National and transnational standards	44
4.1.3 Interoperability	45
4.1.4 Presentation, structure and semantics	45
4.2 Scope of a standard for legislative documents	46
4.2.1 A structural convention	47
4.2.2 A metadata convention and a shared ontology	47
4.2.3 A naming convention	48
4.3 The users of a standard for legislative documents	48
4.3.1 The legislator	48
4.3.2 The legal drafter	49
4.3.3 The toolmaker	49
4.3.4 The citizen	49
4.3.5 The future toolmaker	49

4.4	The tools for the standard-based management of legislative documents	50
4.4.1	The editor	50
4.4.2	The converter	51
4.4.3	Name resolvers	51
4.4.4	Validation tools	52
4.4.5	Document-management and other post-editing tools	52
5	Conclusions (Giovanni Sartor)	54
6	Appendix 1. Managing legislative documents: a review (Enrico Francesconi and Monica Palmirani)	56
6.1	European initiatives	56
6.1.1	Metalex	56
6.1.2	SDU BWB	57
6.1.3	LexDania	57
6.1.4	NormeinRete	57
6.1.5	CHLexML	58
6.1.6	eLaw	59
6.1.7	Legal XML	60
6.1.8	LAMS	61
6.1.9	UKMF	61
6.1.10	The Metalex/CEN initiative	61
6.2	African initiatives	62
6.2.1	AKOMA NTOSO	62
6.3	Asian initiatives	63
6.4	US initiatives	65
6.4.1	An institutional case: XML at the US House of Representatives	65
6.4.2	A non-institutional case: Legal-RDF	66
6.5	Australian initiatives	67
6.5.1	EnAct	67
6.5.2	JSMS	68
7	Appendix 2. Managing of legislative semantics: a review (Mariangela Biasiotti and Enrico Francesconi)	69
7.1	State of the art of emerging technologies	69
7.1.1	Overview of language and knowledge resources: Thesauri, lightweight and foundational ontologies	71
7.2	General Lexical Resources	75
7.2.1	WordNet	75
7.2.2	FrameNet	76
7.2.3	CYC	77
7.3	Lexical resources for legal domains	78
7.3.1	The European Institutional terminological repositories: Eurovoc and Iate	78
7.3.2	The TESEO - TEsauro Senato per l'Organizzazione dei documenti parlamentari	79
7.3.3	The LOIS Project	79
7.3.4	The Syllabus Project	81
7.3.5	The DALOS Project	83
7.3.6	The Euroterm Project (EDC-2214)	84
7.4	Semantic Approach Trends	84
	References	86

1. Introduction: ICT and legislation in the knowledge society

In this section we shall first introduce the present predicaments of legislation and consider how ICT can contribute to address them. Then we shall present legal informatics and legislative informatics.

1.1 ICT and the predicaments of legislation

In the recent years, parliamentary legislation has been criticized for being unable to provide adequate regulatory solutions and for not coping with the needs of our time. We shall argue that ICT can support the mission of legislatures, enabling them to address effectively the present challenges.

1.1.1 Predicaments of legislation

It has been affirmed that the primacy of legislation is coming to an end: we are moving from the age of legislation to the age of administration (by specialized technical bodies), or jurisdiction (creation of law by judges, supported by jurists), or custom (as emerging especially from economic relations, contracts and decisions by private arbiters), or even computer code (enabling or constraining actions in virtual environments). This would be required by a variety of factors:

- Legislative authorities are national; economic and social networks are global;
- Legislation is slow; current problems require quick solutions;
- Elective assemblies are political bodies; complex problems require technical, economical and legal competences;
- Legislation is static; accelerated economical and technical progress requires continuous adjustments.

This picture has undoubtedly some elements of truth, and indeed nowadays parliamentary legislation is just one among other sources of legal regulation, as it appears in Figure 1.

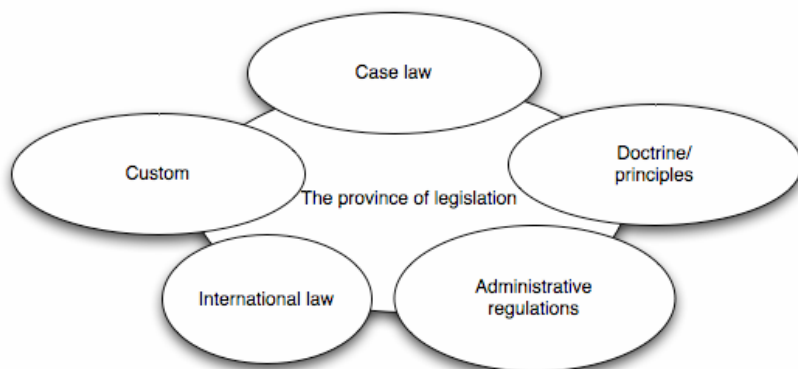


Figure 1: Legislation and other legal sources

The fact that legislation needs to coexist with other sources of the law does not make it less important and less central, it does not diminish the unique contribution that parliamentary legislation can provide to societal governance. Let us shortly consider what we may identify as the values that are inherent in parliamentary legislation, the values which a good legislation should implement:

- Means-end rationality. Legislation should provide regulations that are likely to solve effectively the problems they address, according to the best available knowledge.
- Discursive rationality. Legislation should emerge out of a debate as open as possible, where all interests at stake are considered, the relevant pros and cons are evaluated, and the possible alternative are taken into account.
- Responsiveness to citizens need and preferences. Legislation should reflect the views of citizens, focused on the common good.
- Progress (ability to change). Legislation should be able to adapt to social change, adapting the legal framework to new needs, and, in particular, to introduce comprehensive reforms to accept other sources of law (such as custom or case-law) since it is structurally unable to do.
- Legal certainty. Legislation should contribute to the certainty of the law, providing normative information which gives effective guidance to citizens and legal decision makers. By providing a common focus it should keep in line with their expectations and prevent the exercise of arbitrary power.
- Citizens' rights. Legislation should enable citizens to have a clear idea of what their rights are, so that they may be able to make justified complaints when these rights are violated.

Preserving the dignity of legislation (for this idea, see Waldron 1999) in the information age requires that the values indicated above are approximated, at least to a certain degree, by legislative reality: it is not sufficient to imagine an ideal legislator, it is necessary to show how ideal legislation can become, to a reasonable extent, a concrete reality. The need to improve the quality of legislation is indeed the focus of a number of national and international initiatives¹.

1.1.2 The role of ICT

ICT can contribute to all the purposes that we have indicated at the end of the previous section:

- Means-ends rationality. ICT can contribute to align legislation to its intended purposes by providing legislators with tools for anticipating the impact of new laws on the legal system, on administration, and on society or for monitoring and evaluating impacts of an existing law.
- Discursive rationality. ICT can promote critical debate around legislation, by providing communication tools for promoting the informed debate within parliaments and outside of them, by facilitating the preparation of legislative proposals, and by offering citizens and their associations new ways to participate in the legislative process.

¹ See, for example, OCDE 1994; 1997 and European Commission 2002; 2006)

- Responsiveness to citizens' need and preference. ICT can facilitate the contact between citizens and their representatives, providing citizens with new ways to express their views and with feedback about the choices of their representatives.
- Legal certainty. ICT can enable citizens to anticipate legislation's impact on them, by providing access to laws and cases, by facilitating the drafting of more understandable regulations and the maintenance of the legal systems understandability as much as possible.
- Citizens' rights. ICT can contribute to ensure that laws protect citizens' rights by making more accessible knowledge about rules and remedies and by ensuring publicity of information about officers' behaviour.

We should indeed consider that ICT not only are a part of the problem which legislators currently have to solve, it is also a part of the solution to that problem.

On the one hand ICT are a source of problems for legislators as:

- ICT are the infrastructure of globalization. They enable economic and social networks transcending borders;
- ICT are the engine of economical and social development. They increase the speed of change;
- ICT are the enzyme of complexity: by increasing knowledge and possibilities available to individual actors they increase the complexity of their interaction.

On the other hand, however, ICT are a part of the solution since in particular:

- They can provide an infrastructure for legislative networks in order to enable legislative coordination to address global issues;
- They can support dynamic drafting: ICT enable to control the mass of legislative material, and to intervene upon it making the required changes;
- They can help legislators in modeling laws according to social complexities.

ICT are going to play a fundamental role in the working of parliaments. As in many other domains of social life, also parliaments are now moving into relying upon ICT-based information systems (as opposed to the use of paper as the exclusive mean for storing and communicating data, which could only be processed by humans). The various agents - internal and external to parliaments - that are involved in legislation increasingly access information and communicate mainly through such systems, which become the pivot of their interaction (see Figure 2).

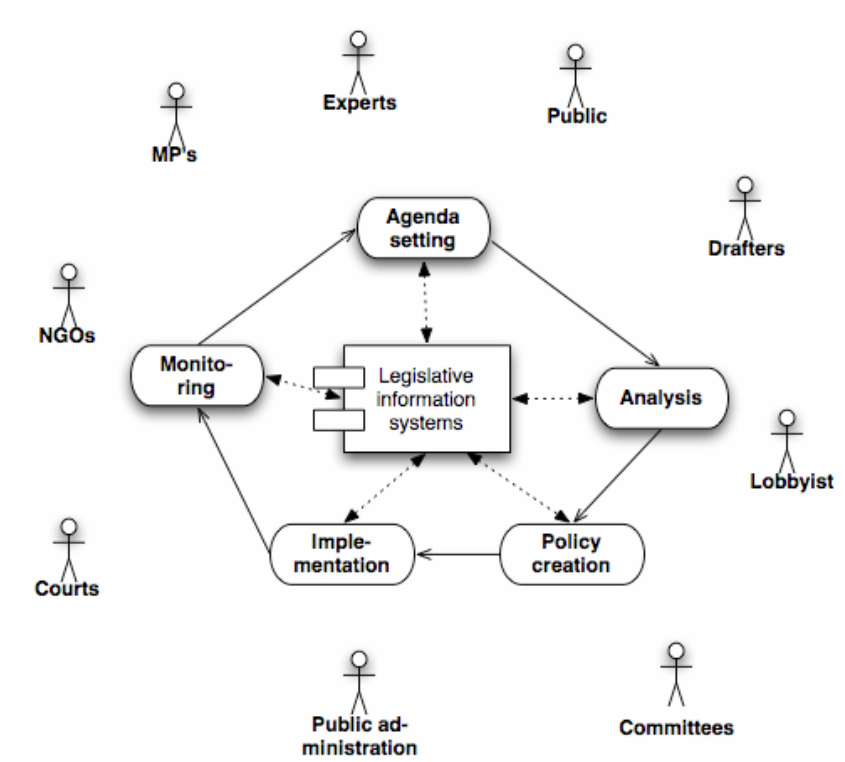


Figure 2: Legislative Information System

Many parliaments are already using advanced information system to support all main internal activities, as well as the connections with the political and social environment (see Figure 3).

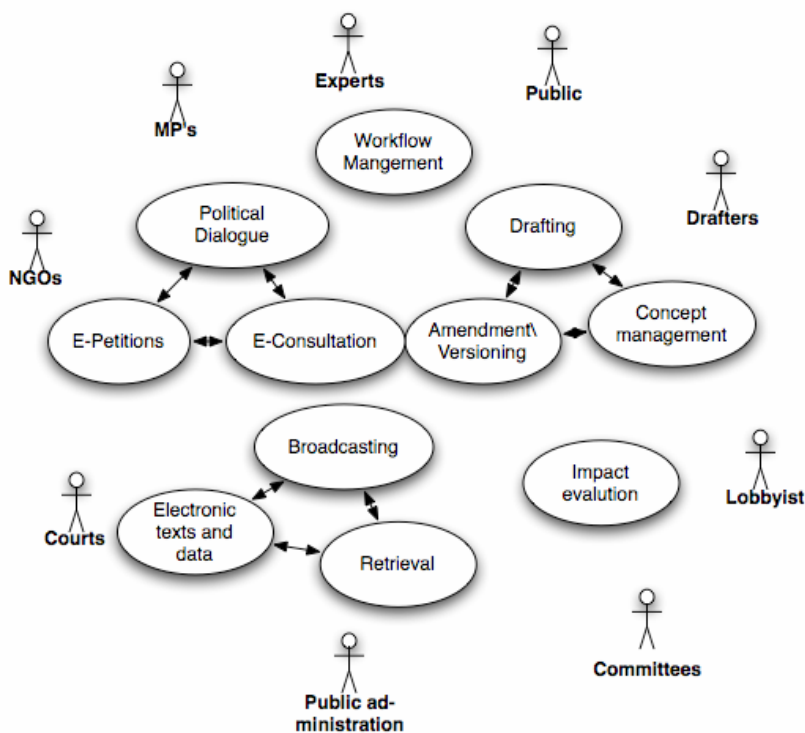


Figure 3: A legislative information system: supported activities

1.1.3 Legislative information systems

The importance of legislative information systems has vastly increased in the last years, in particular since the advent of the Internet, which has dramatically changed the ways all information is produced and communicated (see for all Castells 2000). Nowadays the Internet has already become the principal source of legal cognition for citizens, and it is rapidly becoming the main source of information also for lawyers, as the previously existing legal sources are moving into the Internet, and new sources of legal information are emerging.

The law has massively entered into the so-called knowledge soup, namely, in the mass of information which is nowadays available on the Web: the law constitutes an important set of the information provided on the Web distinguished not only by its particular (legal) content, but also by the fact of being densely interconnected (since legal documents so often refer one to the other, while having relatively few links to non-legal materials). Consequently it is possible to say that a *legal web* is emerging as a distinct subset of the broader worldwide Web. Besides being the source of legal information the Web is increasingly becoming the place of political and legal interactions (for various interesting observations, see among the others, Castells 2001 and Gore 2007, Chapter 9). It is the place where political debate occurs

(consider for instance the many websites and forums currently devoted to US elections), where citizens interact with institutions and with their representatives (through the many institutional sites now provided by public administrations, and in particular by parliaments and other elective assemblies), where legal transactions are performed (through online e-government applications), where legal decisions are commented and criticized (through the many legal fora, blogs, web pages devoted to controversial legal issues, such as the patentability of software).

In the Internet era a parliamentary information system needs to have a Janus-like double face: one face must look at the internal side of the parliament and support all parliamentary activities, the other face must look outside and provide citizens with information on parliament's activities and outputs, as well as opportunities to participate in parliamentary activities.

These two faces need not be sharply divided as far as technologies are concerned: in the Internet era increasingly the same protocols, models, languages, and tools can be used both inside an organization and for its relations with the external world. Following the trend already existing in many domains, each parliament can exploit Internet technologies for building its Intranet, i.e., its private version of the Internet, whose access is limited to MPs and staff, and which offers them the various computer applications available within the parliament, but also new ways of communicating and cooperating. The same technologies and protocols used in the parliamentary Intranet can be employed in the parliamentary Extranet, which gives citizens (as well as their associations and economic and social organizations) access to information and applications existing within the parliament, and ways of participating in the parliamentary activity.

A moving and flexible border separates the information remaining accessible only on the parliamentary Intranet from the information externally distributed through the parliamentary Extranet, a border depending on changing political and organizational choices, as well as on the passage of time (what is internal now, like a proposal discussed in a committee, may and should become public tomorrow, when the proposal is approved and subject to public debate). Thus such information must, from the start, be encoded in such a form as to enable all ways of its future use and distribution (while adopting all technological and organizational measures to maintain its confidentiality, when this is needed and is consistent with transparency requirements).

Moreover, parliaments should not be the only providers of parliamentary information. In pluralistic and democratic societies such information may indeed be distributed—organized and presented in different ways, and possibly enriched with additional information—by other actors, private or public, profit or no-profit (see Section 2.1.6 and Section 3.1.5). Such further activities, though they are not directly controlled by parliament and performed by the parliamentary information system can indeed contribute to the achievement of the main purposes of such system, namely, to support the best use of parliamentary information and to

facilitate citizens' involvement and participation. Such purposes can be indeed achieved not only through direct provision of information services to the internal or external actors, but also by making available to third parties, in appropriate electronic formats, the parliamentary data needed for providing such services.

The link between Parliamentary democracy and ICT is further enhanced by considering that emergence of a global society (constituted by a network of networks of global, national and local social system, based upon ICT-based communication) can be matched by a network of networks of democratic legislative authorities, using a distributed shared set of documents and information. Sharing information not only enables each one to learn from the experience of the other so that best practices can spread, it may also induce national legislators to take into account the needs of others countries, so that they move beyond a narrow view of national interests and frame legal provisions in the light of all concerned interests, also outside national borders (on the ability of democratic process to include the perspective of others, see Goodin 2003).

1.2 Legal and legislative informatics

In the present section we shall shortly introduce legislative informatics, presenting its evolution and its main objects. We shall then delimit the area which is relevant to our concerns, that is, legislative informatics.

1.2.1 A short history of legal informatics

Legal informatics is the discipline which deals with the use of ICT to process legal information and support legal activities, namely, the creation, the cognition and the application of the law.

As automatic processing of information has been expanding from mathematics calculations to data management, office automation, telecommunications and the global knowledge infrastructure, the domain of legal informatics has been consequently expanding.

The first realization of legislative informatics goes back to the 1960's when the first databases of legal documents were started.

The 1970's saw the growth of informatics within the public administration, where large databases were constructed with various administrative data (e.g. population data, fiscal data, etc.).

The 1980's were the time of the personal computer, which enabled the decentralized use of information technology by individual users. This was the era of office automation, when computers entered most legal and administrative offices (both in the public and the private domain), accompanied by suites of user-friendly applications (though the use of information

technologies was generally limited to mundane applications limited to a single user, like text processing, archiving data, using databases).

The 1990's were characterized by the creation of computerized information systems for supporting the activities of many legal organizations, which lead to the integration of the so far separate applications. At this time, in general, automation within legal organizations (parliaments, judiciaries, public administrations) started being conceived as an integrated enterprise. The introduction of ICT within legal bodies ceased to consist in the mere provision of new tools (leaving the existing procedures unchanged); it has become the occasion for reengineering processes according to the potentialities of ICT-based information systems, for rationalizing workflows of enhancing the capabilities of the concerned organization. At the same time legal informatics has started to adopt (and to contribute to) advanced techniques for managing legal information². Analyses of legal reasoning and knowledge developed within legal informatics have proved very useful for legal theory in areas such as legal reasoning (e.g., Gordon 1995, Prakken 1997; Hage 1997, Sartor 2005) and the modeling of norms and normative concepts (e.g., Sartor 2006; 2007). On the other hand models developed within legal informatics have been used within computing and in particular within artificial intelligence, in areas such as belief revision (Alchourryn et al. 1985) or defensible reasoning (Prakken and Sartor 1997).

Between the end of the 1990s and the beginning of the 2000s legal informatics has increased its impact on management and integration of processes concerning production and application of law. Profiting from the opportunities of the Internet, it has expanded its scope to the communications between legal organizations and their public, namely citizens, economic units, and their consultants (lawyers, accountants, etc.). In this way, it has become a significant aspect and an important resource of *e-government*, which in general terms may be viewed as the public sector's use of ICT with the aim of improving information and service delivery, encouraging participation and making government more accountable, transparent and effective.

This is the definition we can find on the UNESCO's website, <http://portal.unesco.org>, which provides the following description of the main sectors of e-government:

- e-administration- refers to improving of government processes and of the internal workings of the public sector with new ICT-executed information processes;
- e-services- refers to improved delivery of public services to citizens (e.g. requests for public documents, requests for legal documents and certificates, issuing permits and licenses);
- e-democracy- implies greater and more active citizen participation and involvement enabled by ICT in the decision-making processes;

As the law is at the core of governance, legal informatics is, therefore, now at the core of e-governance. Its applications span from the legislative domain to the judiciary, to administration, to legal professionals. In all of these areas the use of ICT can contribute to the

² See the biannual International conference on Artificial intelligence and law, ICAIL 2007; or the annual European conference on legal information systems, van Engers 2006.

efficiency of legal processes and to their rationalization, as well as to the realization of legal values such as transparency and controllability (by providing information about legal processes), reasoned deliberation (enhancing the possibility of engaging in informed and reasoned debate), democracy (providing all citizens with information about social problems and their legislative and legal treatment, and new way to discuss such problems and to interact with representative bodies and with their members).

1.2.2 From legal to legislative informatics

Though the scope of legal informatics is not limited to the legislative process or to the provision of legislative information (applications for the judiciary, for administration or for private law practitioners are also important), legislative informatics has a particular significance within legal informatics, as legislation has a prominent importance among the sources of the law.

Legislative informatics achieves its general goal of supporting the legislative process by realizing the following functions:

- Providing information to all actors involved in the legislative process (legislators, citizens, experts, political parties, associations, organizations, lobbies, etc.);
- Enabling cooperation among such actors, so that each one can contribute to the process, according to his or her role;
- Ensuring efficiency, transparency, and control;
- Enabling access to the outcomes of each phase of the legislative process, and contributions to the next phase (managing the workflow);
- Ensuring the quality of the legislative outputs;
- Ensuring knowledge of the produced law texts and preliminary documents.

Though legislative informatics is still is a young discipline, it has achieved in the last years a number of significant results in different areas. To mention a few:

- legislation is now available in electronic format (in many countries),
- it can be accessible online (in many countries),
- the version in force can be automatically constructed (in some advanced approaches),
- retrieval is facilitated by metadata and thesauri or ontologies (in some advanced approaches),
- legislation produced by different bodies and stored in different databases is retrieved through a single Internet-based interface (in some advanced approaches).

Some results are also available in the management of the legislative process (workflow):

- All information concerning a new bill can be preserved and updated during the process (in many countries)
- The legislative process is monitored and supported by workflow facilities (to some extent, in many countries)

- Amendments are updated, so that accurate information is provided on the state of the text as the procedure goes on (in some advanced approaches)

Finally, communication and information support are gaining an increasing significance:

- IT supports interaction between legislators and other actors: between individual MPs and their constituency through e-petition, e-consultation, etc. (in some advanced approaches)
- Relevant information is offered before any legislative choice (in some advanced approaches)

Legislative informatics is also starting to profit from research in the construction of computable models of legislative knowledge:

- Computer models are built of legislative concepts (in some advanced approaches, like the EU project Dalos and Estrella)
- Consistency and completeness of legislation are checked (still experimentally)
- Legislative impacts are being modelled (to some extent, in some advanced approaches, by using different techniques, from traditional statistical analyses to agent-based modelling)
- Knowledge-based systems for the IT-supported application legislation are being developed (in some advanced approaches)³.

As we shall see in Section 3.2.4 a direction in which legislative informatics has realized considerable successes in a very short time is the area of legislative standards (concerning, as we observed above, ways of enriching legal text with machine processable structural or semantic information):

- Standards are available for the structure of law texts (in Europe, Africa and other countries)
- Standards are available for references and modifications (in some advanced approaches)
- Standards are being developed for preliminary materials produced along the legislative process (in some advanced approaches)
- Standard-compliant software is being developed (in many countries)

Using shared standard brings with it considerable advantages, among which there is the possibility of enabling communication and dialogue (vital aspects for deliberative democracy) at all levels:

- improving the quality and accessibility of legal information in different legal systems;
- promoting the interoperability among applications and information systems managing legal information;
- providing high quality integrated services for both policy makers and citizens.

For legislation to cope with the formidable challenge of providing a suitable regulatory framework for the information society it is necessary that legislative authorities are able to make the best use of the many instruments and models provided by legislative informatics. For this purpose, a stricter connection would be desirable between academic work on legislative informatics and research and development taking place within parliaments. Too often academic research does not pay attention to user needs and development initiatives do not pay attention to research (and are not even made known to the scientific community).

³ See for instance Dayal and Johnson 1999

Fortunately, there are recent signs that a more intense cooperation between research in legal informatics and development is underway. On the one hand legislative informatics is producing many results which are usable in legislative practice (for instance, with regard to modelling legal texts, dealing with legal dynamics, building legal ontologies), on the other, development projects in legislative informatics show awareness of academic results. Academic results have been successfully integrated, for instance, in the Italian project Norme-In-Rete⁴, and the United Nations project AKOMA NTOSO⁵, that will be described later.

2 Legislative information and the Web (Giovanni Sartor)

In the present section we will put the issue of the management of parliamentary information in a broader context, namely, in the context of the present trends in the provision of legal information. We shall then focus on the opportunities and challenges related to the emergence of the semantic Web.

2.1 Trends in the provision of legal information

The ways in which legal information is provided have changed enormously in the last years, in connection with the evolution of information technologies. In the following pages we shall identify the main trends that characterize this rapid and still ongoing transformation.

2.1.1 The digitalization of legal information

The first, and most basic trend, is the rapidly progressing digitalization of legal information. An increasing amount of law texts and data is available in electronic formats: legislation, regulations, administrative decisions, case law, contracts, fiscal data, files concerning court proceedings, and so on.

The push towards digitalization is favoured by the integration between the computer-supported drafting of documents and their electronic distribution, and also by the availability of reliable technologies for ensuring authenticity and integrity of electronic documents (like electronic signatures). Electronic legal documents rather than being a copy of pre-existing original paper documents tend to become the official legal texts. The production of private contracts and administrative acts in electronic formats, to be preserved in computerized databases, is today common in most jurisdictions of developed countries.

This trend is being encouraged by the expansion of e-commerce and e-government. A recent development regarding e-government, having major impacts on the law is the document interchange and application interoperability between the Courts' internal users (clerks, judges, etc.) and all their external users (lawyers, expert witnesses, etc.). When this idea will be fully

⁴ <http://www.normeinrete.it/>

⁵ <http://www.akomantoso.org/>

implemented, judicial decisions, in all degrees of jurisdiction, will be in electronic form, and so will be all acts by the parties, the files of each process, the records of the proceedings, all data kept by the clerk's offices. Finally, there are a few experiences concerning electronic promulgation of law text, so that even the original statutory instruments are going to be electronic files, authenticated by an electronic signature of the competent authority (for instance, the Italian Head of State).

2.1.2 The law on the Internet

The second trend, which is synergetic to the first one, consists in the law moving into the Internet. Digitalized legal information, being free from its traditional paper hardware, can be processed by computer and transmitted over computer networks. The Internet is nowadays one of the main sources of legal information for lawyers and citizens. One can find over the world-wide-web many statutory texts, a vast amount of case law, and many comments on laws and cases (there is an emerging vast amount of digitalized legal doctrine). The Web not only is a huge repository of legal information, but is also becoming the open forum where legal issues are debated.

The Internet (in combination with computer nets which are internal to public administrations), moreover, is the place where legally relevant information is exchanged, and where, as a result of such exchanges, legally binding texts are produced. Hence, not only the Internet is the place where one can find information about what legal events are happening in the real world, but it has become a significant component of the legal world, in which many legal events are primarily taking place.

2.1.3 The standardization of formats for legal information

The third trend, which is dependent on the two previous ones, concerns standardization of formats for the representation of textual legal information. As everybody knows, the strength of the Internet is its inclusiveness and openness. One can access whatever information is available and provide (within general legal constraints) whatever information can be uploaded to a connected computer. For the network to grow, a central authority is not required for it is a matter of decentralized decisions of users and providers. However, the available information has to be coded and decoded according to shared machine-readable standards or protocols.

Standards can concern different aspects of legal information, at different levels: the communication protocols that are required for information to be made accessible over the Web; the ways of specifying the typographical appearance of the documents; the links to other documents; the structure of the documents (their division in component units, like sections and subsections); and so on. The determination of machine-processable standards is a crucial issue for public policy in the information society.

On the one hand standardization involves some dangers: (a) standards, once they are established, tend to spread regardless of their merit since the need to participate in communication pushes everyone towards the currently adopted standards; (b) privately

determined standards, or the related algorithms, can be disclosed to others only under specific conditions or can even be the object of intellectual property so that competition can be hindered; (c) the need to respect standards may limit innovation and diversity. On the other hand the shared adoption of appropriate open standard greatly facilitates technological progress, cooperation and competition in the framework of the knowledge society.

2.1.4 The law in the semantic Web

The fourth trend is the move from text-based Web into the so-called semantic Web. This means that legal information available over the Internet is increasingly processed according to its content (or meaning), and not only as a pure text (as a sequence of words). This result is usually achieved by embedding in the natural language text special computer readable specifications, which can be processed in various ways: for retrieving the document, for accessing related information, for determining its currently binding content, for applying the rules it includes, and so on (XML tagging is normally used to include such meta-textual information in legal documents, supplemented with languages like OWL for specifying conceptual structures, and logical extensions like in RULE-XML for capturing the logical structures of legal rules).

Standards for marking derogations and modifications in legal documents are already being devised, which allow the legal text currently in force to be automatically constructed. Further meta-information is being embedded in legal texts available through the Internet for the purpose of conceptual information retrieval. In particular, documents are indexed according to conceptual analyses (ontologies) of the concerned legal domains.

2.1.5 The executable representation of legal information

The fifth trend, just at its beginning, concerns the executable representation of legal regulations. This means that computer systems not only help humans to access legal texts, but they directly apply the legal regulations which are embedded in (or linked to) such texts, or support humans in this task. This is happening in two areas.

Firstly, advanced systems are being developed for automatic contracting, within e-commerce. These systems establish the content of binding contracts for the delivery of goods and services, according to the agreed result of individual negotiation (this takes place particularly in digital-rights-management environments). For such systems to be able to interact meaningfully and correctly both with humans and with other similar systems, it is necessary that they share the same way of representing normative positions (rights and duties) and other legal qualifications.

Secondly, some large-scale rule-based systems have been developed for public administration. Such systems assist employees and citizens in applying legal rules, by performing automatically the corresponding inferences (determining taxes, assessing entitlements, and so on). Both systems for automatic contracting and rule-based systems for public administrations provide proprietary languages for representing normative information, usually not translatable one into the other. The use of such languages (without their interoperability being ensured and

a shared standard being available) may hinder the development of new applications, prevent communication between different systems, and in general reduce competition.

2.1.6 An increasing diversity in information providers

The sixth trend is the increasing diversity in the provision of legal information. In Europe we have seen at first, in the 70's, the emergence of national public systems for the online provision of legal information, while in the US private companies played from the start a major role. Then, in the 80's we have witnessed the crisis of public systems (with a few exceptions), and the increasing provision of electronic legal information by private publishers (using compact disks for distribution). In the 90's public providers have come back, exploiting the fact that law text are available to them in digital format (since they are typeset through computer systems), and that the Internet allows such texts to be distributed at low cost. At the same time, thanks to the Internet, a number of new actors have emerged in the provision of legal information, such as legal information institutes, educational institutions, professional associations, law firms, and others are providing large amount of freely-accessible online legal information (there are even portals which have the specific purpose of providing access to online legal resources).

There is a need to define a new framework for the provision of legal information, which ensures that different information needs (of citizens, public authorities, and professional lawyers) can be satisfied in the best way through the cooperation and competition of different providers. Furthermore, it is urgent to redefine the tasks of public authorities, namely the way in which they should accomplish their duty to ensure knowledge of the law in this complex multi-actor environment (for instance, it has been argued that such authorities do not fulfill their duty by only providing raw texts; they must also structure such texts and enrich them with machine readable meta-textual information, according to the best available standards).

2.2 Focus on the semantic Web

With regard to the management of parliamentary information the move toward the semantic Web (already mentioned in Section 2.1.4) is particularly significant. In the present section we will first shortly introduce the basic idea of the semantic Web and then consider its relevance for legislative information.

2.2.1 The semantic Web

The emergence of the semantic Web results from embedding in the world-wide web (the system of interconnected document today available over the Internet, thanks to the use of the HTTP protocol) information that is machine understandable and can be processed with the available advanced techniques. This represents a significant development with regard to the present situation, where web documents (web pages) mostly contain natural language texts or multimedia files. The promotion of the semantic way is the leading idea of the World Wide Web Consortium⁶ (W3C), whose overarching goal is that of achieving a *one-World-Wide-Web*, shortened into one-Web, that is, an open platform where an increasing amount of

⁶ <http://www.w3.org>

information and services can be made universally available. According to the same consortium, the achievement of the one-Web goal depends on the realization of the following long-term sub-goals:

- Web for everyone, i.e., enabling everyone to go get the benefits of the Web, consisting in the fact that it enables human communication, commerce, and opportunities to share knowledge.
- Web on everything, i.e., enabling every kind of device (including phones, television systems, etc.) to access the Web
- Knowledge base, i.e., enabling the Web to be used not only by humans but also by computer, namely, embedding in it machine processable information
- Trust and confidence, i.e., enabling accountable, secure and confidential transactions.

The main way in which W3C contributed to the one-Web goal consists in providing standards for technologies. Since 1994, they have been producing ‘W3C Recommendations’, which contribute to Web interoperability. This has led in the recent years to the definition of all technological standards that can be seen in Figure 4.

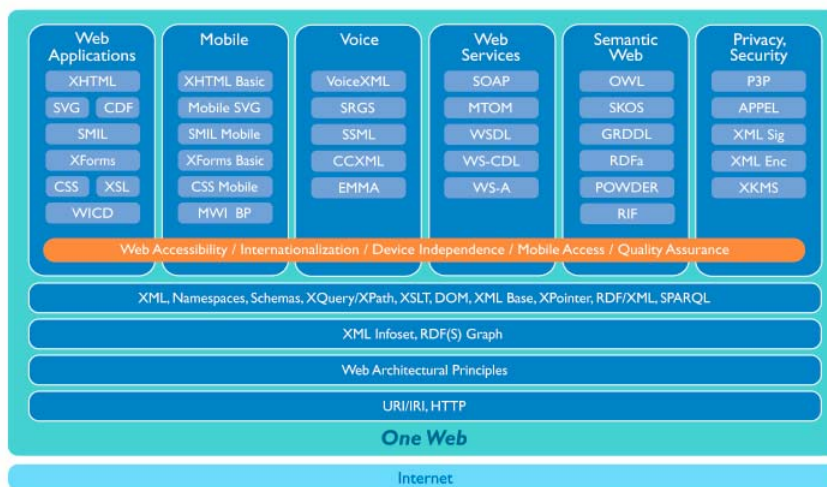


Figure 4: W3C standards

As we have just observed, the W3C, through the knowledge-base aspect of its vision, is leading toward the realization of an emerging reality, the semantic Web, namely, a web where not only information is be made accessible online, but where this is information is increasingly machine-processable. This idea has lead to the design and progressive implementation of the pyramid of technologies you can see in Figure 5, whose layered structure can be described as follows:

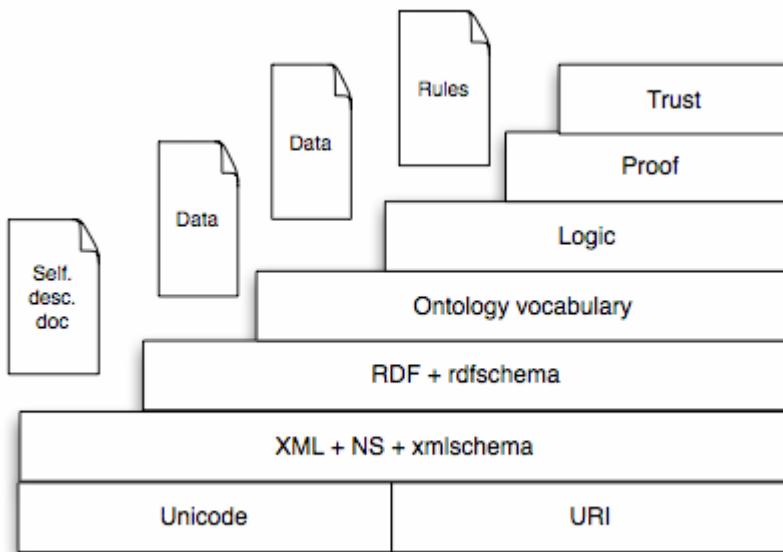


Figure 5: Levels of the semantic Web

- the basic layers is constituted by two standards:
- the Unicode standard, providing an agreed computer representation (a unique number) for any character in any human alphabet.
- the URI (uniform resource identifiers) standard, providing an agreed unambiguous way of naming the entities which are accessible on the Web (the resources).
- XML provides a metalanguage enabling individuals and communities to define tags for expressing the structure of their documents, and for including further information (metadata) to such document, so that the documents can be automatically processed according to such structure and additional information.
- RDF enables us to include in the documents machine understandable statements on relevant objects and their properties.
- Ontologies enable us to specify concepts and conceptual relations and to make conceptual inferences.
- Logic permits to express complex information, so that machines can use it in arguments (proofs).
- The Trust layer includes ways to ensure confidentiality, authenticity, integrity and reliability (such as cryptography and digital signatures).

The realization of the semantic Web will enable us to overcome the limits of the current web, facilitating humans to access information, and enabling various activities over the web (from e-commerce to e-government). On the one hand documents can be automatically adapted to

the needs of their user on the basis of the machine processable information they contain (e.g. a legislative text can be presented in the preferred forms, it can be visualized in the text in force, or in a previous version, etc.); on the other hand documents can embed the information required for e-commerce or e-government, information that can be used by the relevant applications (e.g. names or authors and addressees of a legal act, times for when the legal acts become effective or for deadlines concerning consequential activities, etc.).

2.2.2 The legal semantic Web

Within the emerging reality of the semantic Web we can distinguish a subset of it, which we can call the *legal semantic Web*, which is constituted by the legal contents available in the Web, content that are (or at least may be) enriched with machine processable contents. Thus the "one Web" vision can be translated into a "one legal-Web" vision, namely into a vision where everyone, on any kind of device will be able to obtain reliable legal information, and where legal information will be enriched with machine processable data, so that accessing information and the performing legal transactions are facilitated by computer support.

Indeed we can discern the progressive emergence of such a legal semantic Web, where electronic legal documents are made available on the Web together with machine processable information. The pyramid of web technologies of Figure 5 is indeed being transferred to the legal semantic web, originating the layered pyramid of Figure 6:

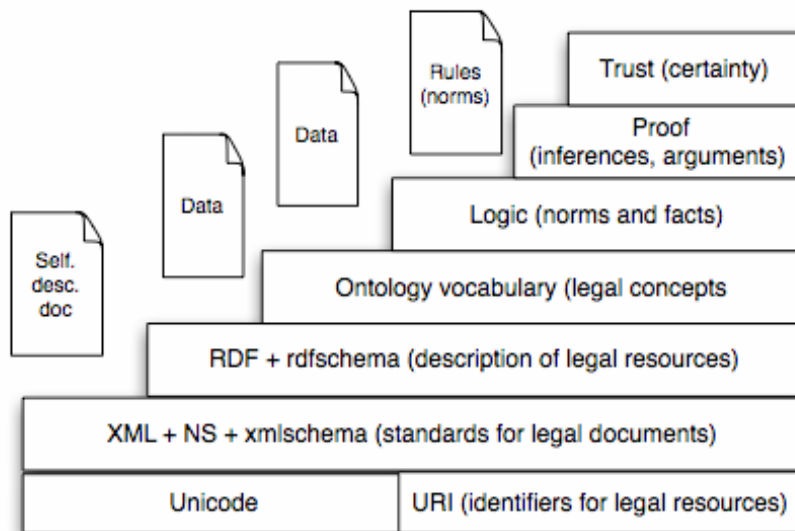


Figure 6: Levels of the legal semantic web

- standards are being defined for identifying legal resources, so that each legal document, produced by any legal authority can be univocally identified (and consequently retrieved);

- standards are being defined for structuring legal documents, of any kind, according to well-specified XML definitions;
- shared standards are being defined for making assertions on legal documents;
- legal ontologies are being created (and linked to general ontologies);
- ways of formally representing legal norms are being devised.

The coherent implementation of such a technological model has so far been accomplished only to a very limited extent (though the trend towards its progressive implementation is already discernible). Thus, even though a huge amount of legislative information is nowadays available on the Web, its utility is limited since this information has the following features:

- it is in different formats (word, pdf, html, xml, etc.);
- it is searchable through search engines (good recall, but little reliability);
- only single pages are retrieved (no integration of data contained in different pages);
- the data within pages is not automatically processable.

Through the realization of legal semantic Web it will be possible to improve substantially the usability of such information by enabling the automatic retrieval of relevant legal data, opportunely selected and integrated, and the automatic processing of legal information embedded in web documents for many different purposes (e.g. generate the law in force, control deadlines, apply rules, etc .)

In the future all actors in the legal world, when looking for legal information will search legal semantic Web, and will be provided with legal information that has been, to some extent, shaped by ICT process (see Figure 7). On the basis of this information they will address the different legal issues in their different roles (as legislators, judges, administrations, parties in economic or other transactions).

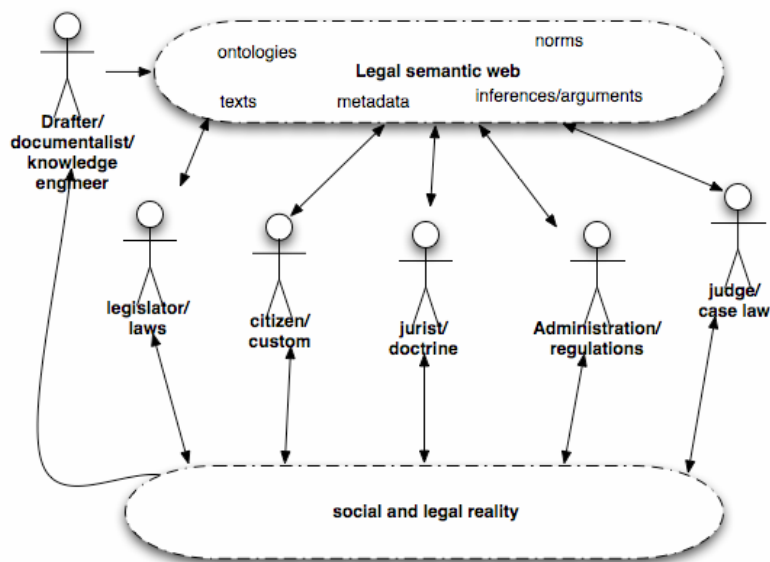


Figure 7: The law between virtuality and reality

The realization of the legal semantic Web, focusing on the addition of computer processable information to legal documents, according to shared standards, facilitates the integration between: a) the production of documents and the management of their workflow and b) the distribution of the resulting documents and information to the public (enhanced with such information). During the production phase, machine processable data can be added to the various documents (there included those having only a preliminary function in the procedure) in order to facilitate the subsequent use of the document and to keep track of its workflow. These same data (for instance the specification of the structure of a document, its authors, its life-cycle, etc.) can enrich the document when they are provided externally, and can be used to provide information concerning the procedure through which the document is produced.

The approach of the semantic Web facilitates on the one hand (on the Intranet side) legal drafting, the maintenance of legal sources and the management of legislative workflows and procedures. On the other hand (on the Extranet side), the publicity of procedures and information, the dialogue between sub-national, national, and international institutions, and the distribution of legal information. The definition of appropriate standards for legislative documents can indeed provide the link between the production of legislation and its use in the community.

3 ICT-based management of legislative documents (Giovanni Sartor)

In this session the focus will be on the specific topic of the present report, that is, the management of legislative documents.

The issue of the management of legislative documents involves both aspects of Parliamentary information systems that were considered above: the “Intranet” aspect, concerning computerization within Parliaments and the “Extranet” aspect, concerning the provision of legal information (and of various services) to external actors.

Moreover, the discussion of the management of legislative document cannot be limited to Parliamentary information systems. As will be discussed below, legislative documents are, and should be, stored and processed also by other actors. Legislative informatics is not limited to the study and development of computer applications for Parliaments. It also includes orchestrating, for the benefit of legislation, activities, applications and systems existing within Parliaments and outside of them (in other branches of public administration, or in the private sector, both profit and no-profit).

This is even more relevant today, in the age of web services (software system designed to support interoperable machine-to-machine interaction over a network, according to the definition in W3C 2004), where complex functions—rather than being executed by an integrated separate system, covering all aspects of them— will be provided by combining the performances of separate heterogeneous systems (agents) interacting according to shared protocols. In particular, as we shall emphasize in the following, the issue of the management of legislative document is not reduced to the issue of their storage and processing within Parliaments, but includes the discussion of how such documents, after being produced by Parliaments, can become shared resources, whose decentralized and autonomous reuse can contribute to the ideal of an open Parliamentary democracy.

The following pages will be devoted to the analysis of the evolution of legal information systems. Firstly, in Section 3.1 we shall describe the achievements of the past, and then, in Section 3.2 we shall consider the currently emerging trends.

3.1 A short history of legal information systems

In the present section we shall shortly describe the evolution of legal information systems in the past century, from its beginnings in the 50’s until the emergence of the Internet in the 90’s, focusing on the issue of documents management (our account mainly relies on Bing -1984- and Bing -2003-), to which we refer the reader for further information⁷. Then we will consider the approaches currently debated, trying to discern an emergent evolutionary trend.

⁷ For a comprehensive account of legal informatics see also Leith and Hoey 1998.

3.1.1 The beginnings

The history of systems related to legal information starts in the second half the 50's when the first computer application in the legal domain was carried out⁸. It was a system developed at the Health law center of the University of Pittsburg, following a request by the legislator of Pennsylvania, to deal with textual amendments. The problem that this legislator had was that of substituting in all State legislation the locution "retarded child" (which was seen as disrespectful) with the locution "exceptional child". Given that performing this task manually would not be only costly but also unreliable, it was decided to transfer the legislation on a magnetic tape (that was the only storing device available at that time), and then search for all occurrences of the string to be substituted. This was successfully done, but once the data were available on an electronic support, it appeared that searching for that particular string was not the only possibility: the tape could be searched for any textual string, to identify and retrieve the piece of legislation containing that textual string. In this way the first instance of electronic text retrieval in the legal domain was accomplished.

This example was soon followed by some entrepreneurial initiatives, aimed at providing computerized information services. At that time, obviously, what could be asked was just a printout of the texts containing the search terms, which were collected by a long search of the tape where the data were collected. No real time answer (and adjustment of the query) was possible.

3.1.2 The thesis: The remotely accessible centralized systems of the 60's and the 70's

During the 60's a number of information retrieval systems for legal sources were created, both in Europe and in North America. In the United States such initiatives were mainly developed by private firms, while in Europe the initiative was mainly taken by public bodies, in particular inside public administration. The motivation behind such initiatives was the idea that legal sources had grown so much in quantity (in particular since the welfare state started to regulate minutely social benefits as well as economic activities) and were changing so quickly that only computers could effectively provide knowledge of the law (this idea was popularized in particular by Simitis 1977). Significant technological developments, taking place between the 60's and the 70's would provide the basis for the successful development of legal information services: the availability of big disk memories, allowing much quicker access than tapes (and consequently enabling real time responses from automatized information systems), and the integration of telecommunications with computing, which enabled remote access to information system (through telephone lines).

In these years in Europe various legal information systems were constructed. For instance in Italy in 1963-4 the Court of Cassation (Corte di Cassazione), the highest judge in matters dealing with civil and criminal cases, started a database of abstracts ("massime") of its own decisions, which grew progressively expanding to the state laws, administrative regulation, etc. (and it is still the most extensive information system available in Italy). Similar initiatives were taken by in France by the Conseil d'Etat (the supreme judge in administrative cases), in

⁸ For a description of the beginning of legal informatics see Chapter 6 - Bing 1984.

Germany by the Ministry of Justice, in Sweden by the Directorat of Court administration, in Finland by the Supreme Administrative Court, in the EU by the legal service of the European Commission. In general the initiative for building such a system would not come from the legislature, but rather from the bodies interested in the application of the law.

Though such systems started with the limited coverage (usually the decisions of a certain kind of courts), they tended to expand with further contents, and in particular with legislation and regulations. The emerging vision was indeed that of a single national legal information system, where all legal sources (legislation, precedents, regulations, treaties, etc.) would be electronically stored and made available for retrieval. This vision underlined the efforts of the 70's and changed under the influence of technological development.

At that time computers were still not used during the phase of document drafting and the original documents were written on paper. Consequently, there was a complete separation between the preparation of legal documents and their subsequent storage in an information computer system (documents would have to be typed again in order to be made electronically available). Thus Parliaments would be mere users of such systems (along with other users). The Parliament staff would make use of electronic information systems during the drafting phase, in order to access the relevant pieces of pre-existing information. This would typically take place through a telephone line, which enabled the remote user to query the electronic legal database (residing in a big and costly computer, a so-called mainframe), to scroll the retrieved documents and possibly to print them. The users of such systems would not be able to reuse the retrieved electronic documents. The printouts would be delivered from a terminal linked to the remote database. As Bing (2003) observes, the only users of such systems were professionals (judges, civil servants, lawyers and their assistants), while citizens could access them only in very limited exceptional cases.

The systems described above had a mixed history of success and failure, and only a few of them survived until our days. In the USA the publicly owned systems FLITE and JURIS died out, while the private providers Westlaw and Lexis have prospered and even expanded to other countries. In particular, Westlaw has become part of Thompson, a huge media company, who has also acquired Lawtel, the legal information system available on Prestel (pre-Internet British system for delivering online legal information). Similarly the independent systems existing in England or Australia did not survive until our days. Conversely, certain systems, such as the Italian service Italgire-Find⁹ and the Norwegian System Noris (managed by Lovdata, a private foundation established by the Ministry of Justice and the University of Oslo) have been continuing until our days, increasing progressively their contents and their services.

3.1.3 The antithesis: The multiple isolated systems of the 1980s

The personal computer (PC) was the invention that characterized the 1980s, leading all of us into the information society. The PC enabled computing to enter all economical and social

⁹ Provided by the Court of Cassation; on the early days of Italgire, see Borruso 1978

activities and also all legal activities, from public administration to the judiciary, legislative, and private law practices. In the legal domain, personal computers started being used mainly for word-processing, but this was soon followed by further applications, such as account-keeping, filing, and information-retrieval. Though PCs could access remote databanks over the telephone lines, the emphasis moved to applications directly available on the PC, whose facility of use and advanced interfaces users had learned to appreciate, and which was not subjected to the costs and the delays of remote telephone service.

Consequently, legal documentation moved to the PC, where it can be independently used by individual users, thanks to the storage of legal databases on a new high capacity storage medium, the Compact Disk, which would be accessed through retrieval software residing on the PC. This technological change determined the entrance of new actors into the domain of electronic legal information like, for example, private publishers who could sell legal databases on CDs through the existing channels available for the sale of book and journals. In comparison with remote databanks, CDs were inferior with regard to the breadth of their coverage and the frequency of update (which required purchasing a new edition of the CD). In fact, many publishers provided CDs merging the various materials at their disposal, namely, official texts and copyrighted doctrinal contributions illustrating the content of the latter. This anticipated the merge of different sources of legal knowledge, which is now taking place in the new framework of the Internet.

One important implication of the office-automation revolution of the 1980s was the computerization of the production of legal texts. At the most modest levels this just involved using computers just as advanced typewriters, but this had the effect that an electronic version of the legal text was available at the very time when the text was produced, without additional costs. Moreover, not only the final document was available in electronic form, but also all preliminary documents (the *travaux préparatoires* in a broad sense) that had been produced during the procedure aimed at the creation and adoption of the final, legally binding, document. This was the case for all various studies, drafts, opinions produced during the legislative process (and similarly for the various documents produced by the parties of a case and by the judge during a judicial procedure).

As a result, computers made possible the integration of the production of legal documents and of their distribution in electronic form. This had different effects, to wit:

- it enlarged the range of documents which could be made available to the public: not only the final, legally binding acts, but also the documents produced at earlier stages;
- it enabled a strict connection between the procedures aimed at producing new law and the legal information systems; the procedure itself could deliver the electronic document, and this document could be used within the same procedure (for its further steps) but could be also made available outside of it;
- it strengthened the role of legal authorities, the producers of the original legal texts, who could start distributing electronically their own outputs, without the need of professional publishers.

Since legal authorities could themselves, with little additional cost, make the legal texts they produced accessible to the public, professional providers of electronic legal information (in particular private publisher) had to change their mission. They could no longer limit themselves to the provision of the bare legal sources, but they needed to provide added value to their legal documents (comments, notes, doctrine, etc.), if they wanted to continue to play a significant (and economically rewarding) role in legal documentation.

3.1.4 The synthesis: The universally accessible but plural systems of the 1990s

The synthesis between the thesis of the 70's (the construction of a unitary National legal information systems, including all legal sources and remotely accessibly by all) and the antithesis of the 1980s (the distribution of multiple instances of multiple databanks, to be in personal computers), was achieved in the 1990s through the Internet revolution, and in particular through the WWW. The Internet is the integration of computing and telecommunications. By making distances irrelevant it enabled the virtual integration of resources resident on different computers; by supporting the seamless interaction of systems and application it enabled merging the services of different applications, interacting with different resources, into complex functionalities; by making systems communicate it enabled communication between their users.

We can say that the Internet provided a synthesis between the unifying approach of the 70's and the separating approach of the 80's, since it allowed each one to access from a single point (name the Internet itself) all legal information, while preserving (and even increasing) the diversity of such information.

One of the major effects of the Internet has indeed been the empowerment of individuals and group. By drastically reducing the costs that are related to the production and public distribution of information the Internet has enabled individuals and groups to contribute actively to the production and the delivery of information, rather than being passive users of it (on this see Benkler 2006).

With regard to legal information, in the past years we have witnessed an impressive, both in quantity and diversity, enrichment of initiatives. As Bing 2003 observes, we have "a profusion of initiatives ...each court, each agency, each institution presented their own site to the public." This increased offer of legal information presents the user with a tendency toward fragmentation. Not only legal information is provided by different agencies and split in different sites, but "these sites often do not have standards for updating response, document design, retrieval strategies etc."

Such puzzling diversity was increased by the fact that not only institutions have been providing legal information, but private law firms too have started to upload legal information into the Web as a way to lure possible clients to their web sites and of informing them on the

competence of the firm. Similarly, non-profit organizations have started to provide online legal information related to their activities¹⁰. Hence, multiple copies of many legislative acts are now available online in a plurality of sites, and in different formats.

The provision of legal information has not been limited only to the reproduction of legal sources but also extended to comments on legal sources (news reports, political evaluations, academic analyses), and more generally, legal doctrine. One of the richest sources of doctrinal information is nowadays the Social Science Research Network¹¹, whose eLibrary consists of two parts: an Abstract Database containing abstracts on over 162,600 scholarly working papers and forthcoming papers and an Electronic Paper Collection currently containing about 130,000 downloadable full text documents in Adobe Acrobat pdf format, a large section of which is constituted by papers in legal doctrine.

In addition, the domain of legal blogs should also be considered. Blogs rapidly became one of the richest sources of doctrinal information¹². There are also many discussion groups devoted to legal issues and even websites devoted to the presentation-discussion-advocacy of important legal topics¹³.

The Internet's synthesis has led to an enormous growth of the body of legal information available to all—the shared *Jurisphere*—and this expansion has benefited legal researchers and many practitioners. However, it is not easy task to find the information one needs in the Jurisphere as such information has to be extracted from a huge and diverse collection of material, dispersed in multiple sites, with different standard of document design, retrieval strategy, level of accuracy, update strategy. Typically we use one of the generally available search engines and find all texts containing certain words or combinations of words, ordered according to their relevance (on the basis of the relevance algorithm used by the search engine). The output can be a big set of documents, but such a result might just represent the beginning of an inquiry and one would have to browse through the retrieved documents, checking their pertinence with regard to the objective, reliability of their source.

Even after solving the problem of noise (false positives, namely the retrieval of irrelevant documents), we would have to solve the problem of silence (false negatives, namely, failure to retrieve relevant document), which is only partially addressed by the richness of the documentation available. In fact the documents that can be automatically retrieved remain limited to those containing the words in the user-query. For addressing this problem we would need an advanced search, able to map the user-query to the documents that, using whatever linguist form, express information relevant to the user. Moreover, some further documents are

¹⁰ see for instance the various legal sources on human rights at <http://web.amnesty.org/>

¹¹ <http://www.ssrn.com>

¹² See for instance <http://www.becker-posner-blog.com> where two leading legal academics address current legal and political issues

¹³ See for instance <http://www.nosoftwarepatents.com/en/m/intro/index.html> on software patents

relevant by their relations to documents pertaining to the users query (consider for instance laws modifying, abrogating or suspending the documents which deal with the users' interests).

Finally, even when the users retrieve the document most relevant to their interests, unless they are assisted by advanced legal information systems, they still have the task of combining the different bits of information (possibly contained in different documents) pertinent to their problem (for instance, they would have to construct the law in force by mentally performing all the required textual modifications).

3.1.5 Legal information institutes

The Internet, besides enabling anyone to add his or her bit of information to the available “legal-knowledge soup”—by republishing existing sources (possibly in new combinations), producing new versions of existing sources (for instance consolidated texts), or adding original comments or doctrinal contributions—has also enabled new actors to engage in the construction of broadly scoped legal information systems. This is the case in particular for Legal information institutes (LSI), namely, independent bodies aimed at providing free access on a non-profit basis to multiple sources of legal information, originating from multiple public bodies¹⁴.

The activity of such institutes is based upon their idea that all legal limitations to the duplication of legal information (as deriving in particular from the so called Crown Copyright which State entities traditionally have in Commonwealth countries over official documents), should be overcome, as it is affirmed in their joint “Declaration on Free Access to Law” (see Free Access to Law Movement 2004), which states the following principles:

- Public legal information from all countries and international institutions is part of the common heritage of humanity. Maximizing access to this information promotes justice and the rule of law.
- Public legal information is digital common property and should be accessible to all on a non-profit basis and free of charge.
- Independent non-profit organizations have the right to publish public legal information and the government bodies that create or control that information should provide access to it so that it can be published.

According to the same declaration, public legal information is defined as follows:

Public legal information means legal information produced by public bodies that have a duty to produce law and make it public. It includes primary sources of law, such as legislation, case law and treaties, as well as various secondary (interpretative) public sources, such as reports on preparatory work and law reform, and resulting from boards of inquiry. It also includes legal documents created as a result of public funding.

¹⁴ See <http://www.worldlii.org>

The experience of the Legal information institutes started in 1992, when the Legal Information Institute of Cornell Law School¹⁵ was launched by Peter Martin and Tom Bruce. One aspect of vision of Legal research institutes, which strongly distinguished it from all previous initiative in legal documentation, consists in extending electronic access to the law also outside professional user. As Peter Martin puts it:

One of our powerful early discoveries was how much demand outside those professional sectors there was, ordinary citizens trying to make sense of laws that impinge on their lives (cited from Bing 2003).

Cornell's initiative (which covers now codes, federal laws, decisions by the Supreme Court and by the New York Court of Appeal and many further collections of legal materials) was followed by the creation of other similar institutes in many countries. In particular, the Australasian legal information institute¹⁶ has been particularly successful (Greenleaf et al. 2002) and has represented the model for further initiative in various countries, like Canada, UK and South Africa. A World Legal Information Institute has also been established, which provides an access point to the collections of various Legal Information Institutes, containing 865 databases from 123 countries and territories.

Bing (2003) observes that legal information institutes provide a useful service in contributing to make original legal sources available to both legal professionals and lay citizens, as required by the idea of *publicatio legis* (the idea that the law should become binding only after being made available to all) an idea which is related to fundamental principles such as those of legal certainty and democracy. However, according to Bing, such institutes have now to face two problems. Firstly it may be doubted that knowledge of the law can be effectively ensured by providing citizens with the original texts. Lay citizens would rather need "a problem-oriented gateway to the material, where the authentic instruments are commented and explained". Secondly, as we have observed above, the function of providing the original sources tends to be assumed directly by public bodies. Such bodies can build today their electronic repositories as a (almost) costless byproduct using ICT for the purposes of drafting and workflow-management, which enables them to provide free generalized access to their sources without the mediation of a third party.

3.2 The dialectics of current approaches

Different trends and different models can be anticipated for legal information systems in the framework of the Internet synthesis we described in Section 3.1.4. We will distinguish two main approaches, again a thesis and an antithesis:

¹⁵ <http://www.law.cornell.edu>

¹⁶ <http://www.austlii.edu.au>

- the thesis (presented in the next section) consist in a newly-conceived centralized and comprehensive public legal-information system, integrated with the management of legislative (and judicial) procedures, and using the Internet for dissemination;
- the antithesis (presented in the Section 3.2.2) consists in the decentralized non-governmental development of legal information systems and services on the basis of the public availability of electronic legal documents.

We shall than argue (in section Section 3.2.3) that a synthesis of these two model is currently emerging, where independently prepared and managed legislative information is unified through the use of shared standards.

3.2.1 A new thesis: A centralized legal information system for the Internet age

This thesis is clearly presented by Jon Bing (Bing and Schoenberg 1994, Bing 2003), who argues that the way ahead in legal documentations is represented by the move from retrieval into *regulatory management*. This requires adopting a different view of the services to be provided legal information systems, a view that is not restricted to their being used for conducting legal research (for retrieving stored documents), but which also includes their use in the preparation of new regulations (drafting, reviewing, amending, etc.) and in the application of such regulations (by judges and administration).

This approach, he argues, is mostly appropriate for developing countries, which are not — like the North-American or European jurisdictions — “seeped in traditions and established arrangements, where the new computerized service has to find its place among legal publishers of primary and secondary sources, legal gazettes, and other well established practices.” Such an integrated system should cater for the need of different users: the legislator, the administration, the judiciary and private lawyers, as well as the common citizen (to the extent in which the latter may be content with the legal sources).

According to Bing’s vision, the core of such a system should consist of a mother-database, containing all documents, marked according to the same markup language.¹⁷ Such a database should be prepared and updated though a centralized editorial process which controls the input documents, normalizes them, and gives them the appropriate format (see Figure 8).

¹⁷Bing considers a markup based on SGML (Standard Generalized Markup Language), a metalanguage for defining markup languages for documents. Now SGML has been almost completely substituted by XML, a simplified rework of it.

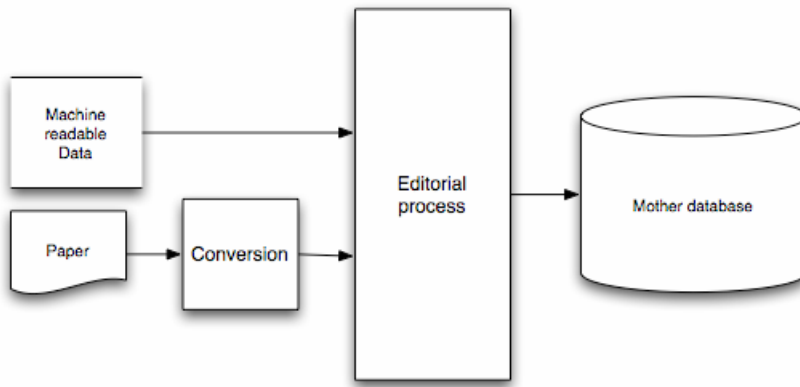


Figure 8: Jon Bing: the production of a mother-database

From the mother-database different outputs in different format should be produced (see Figure 9).

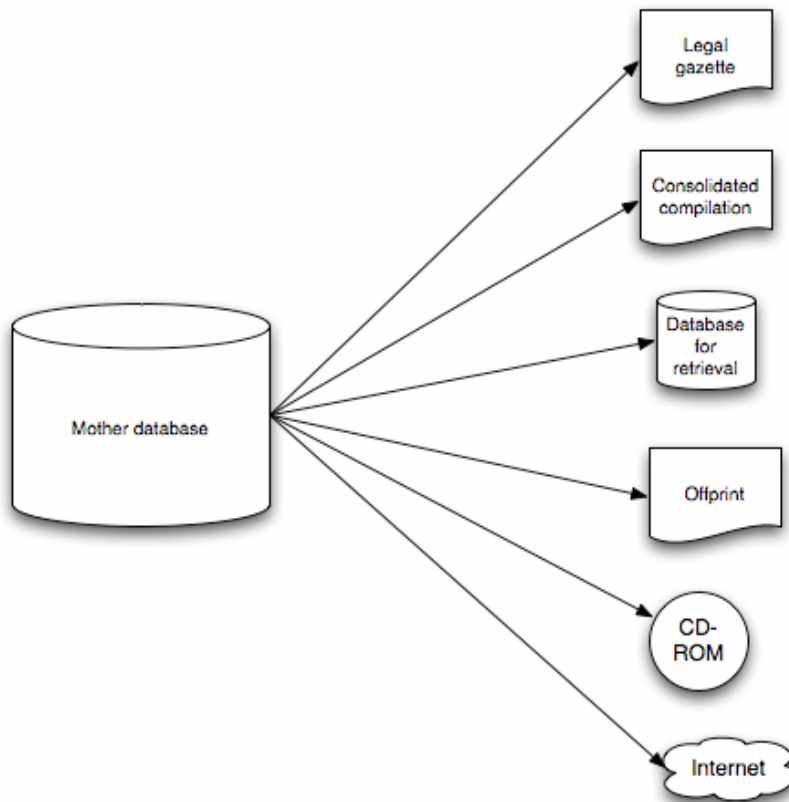


Figure 9: Jon Bing: the outputs of the mother-database

In particular, the mother-database should deliver the following outputs: a *status* database (a database that provides the law in force, and offers high performances to professional users), a legal gazette with the new legislation, compilations of the law in force, and freely accessible text for Internet browsing.

Note that in many jurisdictions a gazette in electronic form is already published together with the gazette in paper form. Given that electronic text can be delivered immediately and costless, the electronic gazette is gradually substituting the paper version as the main reference for professional lawyers and for ordinary citizens. This fact has been officially recognized in some countries, like Norway or Austria, where the electronic publication already represents the original legal text. In Austria, in particular, the federal Constitution has been changed so that the text published online in the Austrian Legal Information systems constitutes the official gazette¹⁸.

3.2.2 A new antithesis: Access to distributed legal resources

The antithesis to Bing's centralized solution (universal access to a central official legal database) consists providing one or more universal access point to distributed legal resources. Such resources will reside in the information systems of the authorities creating the stored legal texts, but different content providers (public or private, profit or non-profit) will provide access to them, either just by building an index to such resources or by also copying them to a central repository. In this model the Internet provides both the channels through which information is extracted from the distributed original databases and the channel through which the user can access such information, as you can see in Figure 10.

¹⁸ <http://ris1.bka.gv.at/authentic/index.aspx>.

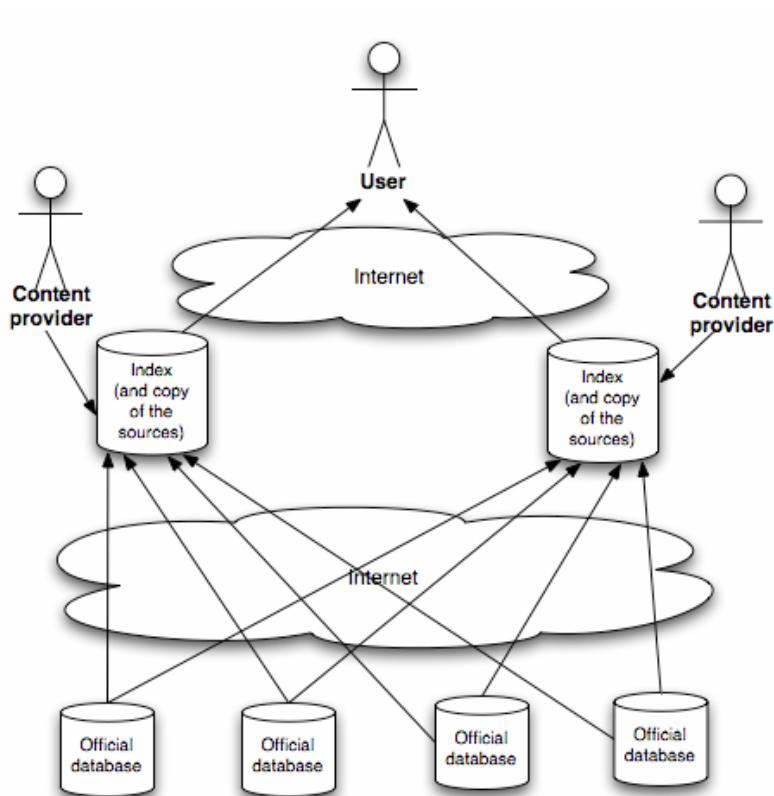


Figure 10: Providing information from distributed legal databases

The latter view is advocated by Graham Greenleaf, one of the leader of the Legal Information Institutes movements, according to whom the centralized solution proposed by Bing may be adequate to certain context (small countries having an efficient, homogeneous and integrated public systems, like the Scandinavian ones), but it is not appropriate for coping with large, heterogeneous and diversified contexts. For legal systems fitting within the latter profile (and for integrating different legal systems), only a decentralized architecture can work.

Greenleaf argues that for achieving adequate provision of legal information it is not necessary that public authorities directly provide a service covering all possible needs: such needs can also be covered in different ways, by different providers, among which an important role can be vested by the Legal Information Institutes. Public authorities are rather required to comply with the obligation to provide “full free access to the law” by which he means the obligation of the legislator (as on any other producer of legally binding documents) to provide “free access to the computerized sources of legal data to those who wish to publish it”(Greenleaf, 2004). More exactly, according to the philosophy of the Legal Information Institutes, legal authorities have the obligation to make available legal information in such a ways that its provision satisfies the following requirements:

- Provision in a completed form, including additional information best provided at source, such as the consolidation of legislation, and the addition of catch words (index terms) or even summaries to cases.
- Provision in an authoritative form, such as use or court-designated citations for cases and (eventually) use of digital signature to authenticate the versions distributed.
- Provision in the form best facilitating dissemination, which should always now mean in electronic form, should in most cases be possible by e-mail or more sophisticated form of data delivery, and should be possible in a form facilitating conversion.
- Provision on a marginal-cost-recovery basis to anyone, so that governments do not attempt to profit from the sale of public legal information, thereby creating artificial barriers to access to law.
- Provision with no reuse restrictions or license fees, subject only to such minimal restrictions as are necessary to preserve the integrity of published data.
- Preservation of a copy in the care of the public authority, so that an archive of the data is preserved to enable greater competition whenever a new entrant wishes to publish the data and whether or not the public authority publishes the data itself.
- Non-discriminatory recognition of citations, so that Court-designated citations are not removed from “reported” cases, ending the privileged status of citations of “official” reports. (Greenleaf, 2004, 69)

Once public authorities have made accessible legal text in this way, they can rely on non-governmental actors for distributing and enriching legal information (or can intervene by supporting the operation of such actors, rather than directly).

Greenleaf also observes that nowadays the task of collecting different sources into a distributed unique database does not necessarily require an editorial process, but can be performed automatically by spiders or web robot (also called crawlers) which can peruse the Web indexing all relevant sites, and that the noise in information retrieval can be reduced by using relevance ranking as performed in leading search engines such as Google.

3.2.3 A new synthesis: Standard-based legal information in the jurisphere

Both approaches just described present significant advantages. The model of the unique official mother-database is based upon the idea that all legal sources should be stored in a uniform format, according to a consistent editing procedure, and that such sources should include all machine-processable information needed to support subsequent multiple uses. Consequently, it seems capable of providing significant advantages:

- Reliability of (and consequently trust on) legal information would be ensured, since such information would always be extracted from the official database.

- Formal coherence of the different legal sources and of their formats would be achieved (for instance, in the way of expressing references), since all of them would result from the same repository.
- Noise in information retrieval would be reduced, since only the official database could be searched (rather than the multiple overlapping repositories now available over the Internet).
- No strong technological requirement would be put on the drafters of legal sources, who could just use ordinary word-processing tools, since the texts would be structured and enriched with metadata only at the moment of their transfer into the mother-database.

Also the idea of extracting information from distributed legal database—and more generally, from distributed repositories of legal sources, possibly managed by the same authorities who have adopted the stored normative acts—has distinctive advantages:

- It would enable a diversified and competitive provision of legal documents.
- It would facilitate the integration between legal procedures and the provision of legal information (each producer of normative act would deliver to the public the documents in the same electronic form they had when resulting from the procedure leading to their adoption).
- It would provide decentralization and autonomy, since each (kind of) authority would be able to organize its document management-system according to its needs, possibly enriching it with additional services (e.g. a point-in-time legislative database), and tailoring it to the needs of its users.

Fortunately, we are not facing here a tragic choice between the two approaches just presented, that is, a choice where the adoption of one approach would entail losing the advantages of the other. In the last years a synthesis has emerged that is able to preserve the advantages of both approaches. This synthesis is based on the decentralized production of electronic legal (and in particular, legislative) document according to shared standards. Thus, the unifying aspect no longer consists in the creation of a unique database or in a uniform editorial activity, but rather in the adoption of a common standard specifying how legislative and other legal document are to be given unique names and how they can be enriched with machine processable data specifying their structure, indicating their links, and describing their content. The standard should provide ways of expressing non only the structural elements of the text (articles/sections, subsection, chapters, titles, etc.) but also references and modifications (so that, in particular, the text of the law in force can be automatically constructed).¹⁹ Moreover the standard should be extensible so that further components can be accommodated in it, if required by further applications (e.g. for checking the logical consistency of the norms expressed by the text, for analysing the ontology of the concepts it uses, and so on).

According to this vision, standard-compliant legal documents, produced by different entities, but primarily by the bodies adopting such documents, can be stored in distributed databases or in centralized ones, can be drafted and processed using the same standard-compliant software, can be reused by any person or system knowledgeable of the same standard, can be enriched

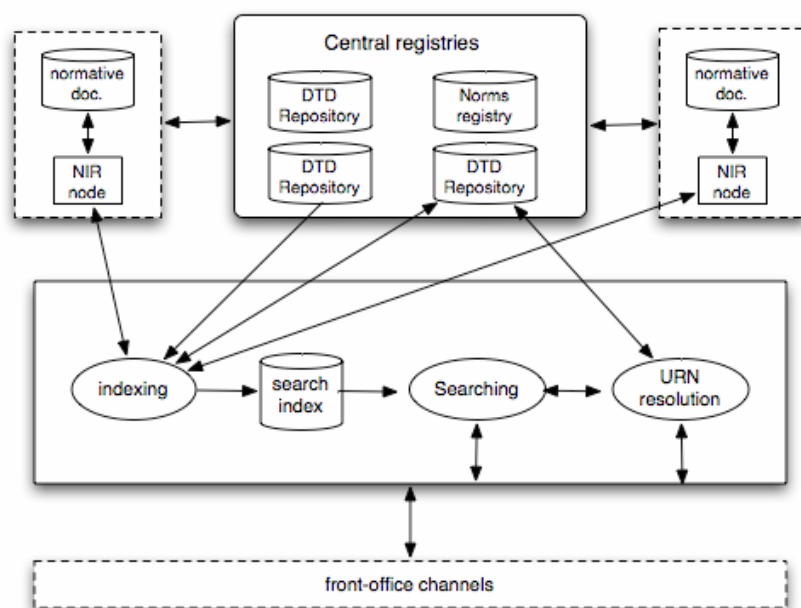
¹⁹ See Palmirani and Brighi 2006. For a logical analysis see Governatori et al. 2007.

by adding further information and ways of processing it. In this way that the interested users could be provided with new derived legislative works, with could be constructed by the Parliamentary offices or by third parties using the legislative materials. For instance, texts could be expanded with doctrinal comments or semantic metadata, or even with a machine-processable representation of concepts and norms which could then be applied with the support of a knowledge-based system.²⁰ However, the official instances of such documents will remain distinguishable and their content (both the text and the originally added machine-processable information) will be controllable and reliable.

3.2.4 Standard-based management of legislative documents: The Norme-in-rete example

One significant (though still at an initial stage) example of what can be achieved by a standard-based approach to the management of legislative documents is constituted by the Italian project Norme in Rete. On the basis of the definition a common standard for legislative documents, a federated system has been developed, based on a distributed model involving all bodies which adopt normative acts: Parliament, Government and ministries, authorities, local autonomies.

In this model each public administration is supposed to store its document in a separate database, but to structure such documents according to the shared standard, and to make them accessible to centralized retrieval facilities. The central indexes are automatically built by web spiders visiting the sites of the federated authorities (Figure 11 represents the architecture of the system).



²⁰ See for instance Dayal and Johnson 1999; For a general discussion on the formal representation of legal contents see Sartor 2005

Figure 11: The cooperative architecture of “Norme in Rete” (adapted from slides by Caterina Lupo)

Note that in a first phase of the project the involvement of the Italian Parliament has been limited: Italian laws, when approved by both chambers, are promulgated by President of the Republic, and then published under the responsibility of the Minister of Justice. The final text is provided by the database of the Court of Cassation (managed by the Ministry of Justice), as well as by the official journal. In the future, however, a stronger Parliamentary involvement is being considered:

- the text of the Acts approved by the Parliament (and transmitted to the President of the republic for promulgation) must already contain all structural information and the metadata that are available at that stage (obviously data about subsequent activities, such as the promulgation of the document can only be provided after such activities have taken place).
- the Parliament should make available to the public all documents produced during the procedure leading to the approval of legislative acts.

Note that in such a model no editorial intervention is considered, since it is assumed that new documents will be provided by the normative institutions in a standard-compliant format, and that all software dealing with such documents (for drafting them, for managing their workflow, for their storage, for their retrieval, for their further processing) will take the standard into account. It is assumed, on the contrary, that exactly the availability of a common standard will be a decisive factor favouring the development of software tools enabling that documents are prepared from the start in a standard-compliant form. Figure 12 shows the expected final outcome of the Norme in Rete project.

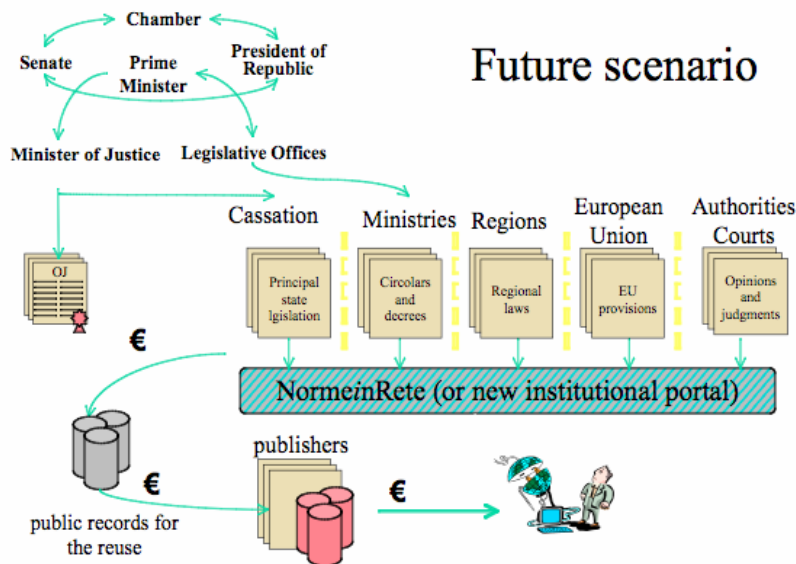


Figure 12: The "Norme in Rete" distributed model

Currently, besides setting up this new framework for new legislation, existing legal documents are edited, with the support of automated tools, in order to set them in the appropriate format and make them available in standard-compliant databases.

Interestingly, the project is committed not only to providing information to the public, but also to making normative documents, in the required format, available to publishers and other third parties so that such documents can be reused and further distributed. According to this idea there is no state monopoly on legal information; on the contrary, the provision of legal information to the public is open to the contribution of non-governmental profit and non-profit institutions (such institution indeed, while aiming at their legitimate commercial or non-commercial purposes also contribute to the public objective of increasing the knowledge of the law). This is indeed what is required by the European discipline on reuse of public information of Directive 2003/98/EC (EU, 2003), which requires (art. 5) that “Public sector bodies shall make their documents available in any pre-existing format or language, through electronic means where possible and appropriate” and that “Where charges are made, the total income from supplying and allowing reuse of documents shall not exceed the cost of collection, production, reproduction and dissemination”.

As a matter of fact, a double channel to publicly distribute legal documents is emerging in Italy:

- on the one hand the Norme-in-Rete portal offers central access to the distributed databases of the different authorities;
- on the other hand all such documents are reproduced in the central database held at the Italian Corte di Cassazione, the system we mentioned in Section 3.1.2.

The latter database has in fact evolved into accepting the Norme-in-Rete standards. This is respected by the new documents being inputted into the Cassazione-database, and old documents are being reformatted (and supplemented with the required meta-information) so that they comply with Norme-in-Rete. In fact, in the Internet framework (where the physical location of digital objects is irrelevant, being all of them instantaneously accessible) it does not really matter whether legal documents are (only) retrieved from distributed databases or whether they are retrieved (alternatively or additionally) from a central databank. What matters is that the official version of such documents is provided by the competent authority and that such texts are made available for retrieval from a central facility, as well as for further processing by any entities that may be interested in doing so.

3.2.5 Standard-based management of legislative documents: The AKOMA-NTOSO example

The Norme-in-rete project is just one instance of the many projects centred upon standards for legislative documents. Other countries have adopted similar projects. Here we just need to observe that besides national projects, some international projects aimed at standardizing the format of legal documents (while respecting national traditions concerning the drafting and the

presentation of such documents) have been undertaken. By providing internationally agreed standards such projects contribute to results that would not be obtained on the basis of merely national standards, namely, supporting the exchange of legal information at a transnational level, and creating competitive transnational market for standard-based software products.

Among the international projects, we can mention the AKOMA-NTOSO project (Vitali and Zeni 2007), which has been developing standards for legislative documents in Africa. The project advocates the use of such standards with a legal drafting methodology and with the provisions of a system for supporting legislative activities (Bungeni). AKOMA-NTOSO does not provide only an XML standard for legislative acts, it also provides standards for documentations, legislative reports, debate reports (Hansard). Moreover, a standard for judicial precedents is also being defined. Here is how the AKOMA-NTOSO model is described²¹:

The AKOMA-NTOSO model has been informed by the following strategic goals:

- To create a “lingua franca” for the interchange of Parliamentary, legislative and judiciary documents between institutions in Africa. For example, Parliament/court *X* should be able to easily import a piece of legislation made available in AKOMA-NTOSO format by Parliament/court *Y*. The goal here is to speed up the process of drafting new legislation/writing sentences/etc. by reducing the amount of re-keying, re-formatting etc. required.
- To provide a long term storage and access format to Parliamentary, legislative and judiciary documents that allow search, interpretation and visualization of such documents several years from now, even in the absence of the specific applications and technologies that were used to generate them.
- To provide an implementable baseline for Parliamentary, legislative and judiciary systems in African institutions. It is envisaged that this will lead to one or more systems that provide a base layer of software “out of the box” that can then be customized to local needs. The goals here are twofold. Firstly, to facilitate the process of introducing IT into African institutions. Secondly, to reduce the amount of re-invention of the wheel that would result if all institutions pursued separate IT initiatives in the area of Parliamentary, legislative and judiciary document production and management.
- To create a common data and metadata models so that information retrieval tools & techniques used in Parliament/court *X* can be also be used in Parliament/court *Y*. To take a simple example, it should be possible to search across the document repositories of multiple Parliaments/courts in a consistent and effective way.
- To create common resource naming and resource linking models so that documents produced by Parliaments/Courts can be easily cited and cross-referenced either by other Parliaments/courts or by other users.
- To be “self-explanatory”, that is to be able to provide all information for their use and meaning through a simple examination, even without the aid of specialized software.

²¹ <http://www.akomantoso.org>

- To be “extensible”, that is it must be possible to allow local customizations to the models within the AKOMA-NTOSO framework so that local customization can be achieved without sacrificing interoperability with other systems.

Another international project deserving particular attention is Metalex (Boer et al. 2002), which, differently from AKOMA-NTOSO , does not aim at directly providing a standard for legislative document, but aims instead at providing a way of mapping different standards, so as to support the interchange of legislative materials. Metalex has been integrated in the framework of the Estrella project with LKIF, a proposed standard for dealing with legal contents, i.e. norms and ontologies (see Gordon 2007, Boer et al. 2007).

4 A standard-based approach for the management of legislative documents (Fabio Vitali)

This section will present the emerging standard-based approach to the management of legislative documents, an approach that best fits the needs of a Parliamentary democracy in the Internet age.

As we have already observed, this approach focuses on a definition of the ways in which machine readable information is to be added to the official text of legislative documents. On the basis of this definition two kinds of software tools can be developed: tools supporting the preparation of standard-compliant documents (by helping the drafter to embed the machine-processable information in the document) and tools managing the documents compliant with standard (on the basis of the machine-processable information embedded in them). In this way a seamless integration between the production of legislative documents and their subsequent use is enabled.

For *open access* to be ensured, such a standard should be *open* in multiple senses:

- its use should not be limited by intellectual property rights²² ;
- its definition should be publicly accessible and understandable to all (having the required technical competence), being expressed in a publicly known format (metalanguage);
- the machine-processable information added to the document should also be accessible and understandable ;
- the standard should allow extensions and adaptations to match specific present and future needs (extensibility);
- the development of the standard should be entrusted to a body open to the participation of all interests on which the standard impacts and on all capacities which are useful for its development.

In the following sections we will first present the purpose and the scope of a standard for legislative documents, and then we will identify the user of standard-compliant documents, and finally we shall consider the different software tools needed for a standard-based management of legislative documents. Our model will be strongly based on the ideas characterizing the project AKOMA-NTOSO (as described in the Africa i-Parliaments Action Plan of the United Nations Department of Economic and Social Affairs), which, among the initiatives so far developed, mostly embodies the principles of a standard-based management of legislative documents.

4.1 The objective of a standard for legal documents

Good standards for legislative documents define a set of simple, technology-neutral representations of Parliamentary documents for e-Parliament services and provide an enabling

²² See the W3C patent policy on <http://www.w3.org/Consortium/Patent-Policy-2004\0205>

framework for the effective exchange of “machine readable” Parliamentary documents such as legislation, debate record, minutes, etc.

Providing access to primary legal materials and Parliamentary documents is not just a matter of giving physical or online access to them. As we have observed above, “Open access” requires the information to be described and classified in a uniform and organized way so that content is structured into meaningful elements that can be read/understood by software applications, and the content is made “machine readable”.

4.1.1 A rationale for shared standards

A standard for Parliamentary documents contributes to fulfill the citizens’ right to access Parliamentary proceedings and legislation by providing “open access” and advanced functionalities like “point-in-time” legislation through standardized representations of data and metadata in the Parliamentary domain and mechanisms for citations and cross referencing of legal documents.

Regardless of the processes that generate and use Parliamentary documents, regardless of the cultural and historical factors that give shape and substance to Parliamentary documents, and regardless of the human languages in which these documents are written, there are undeniable relationships that connect documents of the same and different types. There are also similarities among documents of different countries. One of the main objectives of a standard for legislative documents is to be able to capture and describe these similarities so as to unify and streamline, whenever possible and as far as possible, the processes and formats and tools related to Parliamentary documentation. This lends itself to reducing investments in tools and systems, helping open access, and enhancing cooperation and integration of governmental bodies both within the individual countries and between them. A standard would define a model for open access focused on the following issues:

- generation of documents: it should be possible to use the same tools for creating the documents, regardless of the type, country, language, and generation process of the document.
- presentation of documents: it should be possible to use the same tools to show on screen and print on paper all documents, regardless of their type, country, language and generation process.
- accessibility of documents: it should be possible to reference and access documents across types, languages, countries, etc., implementing the network of explicit references among texts into a web of hypertext links that allow the reader to navigate easily and immediately across them.
- description of documents: it should be possible to describe all documents, regardless of their types, languages, countries, etc., so as to make it possible to create repositories, search engines, analysis tools, comparison tools, etc.

At the same time, a good model considers the differences that exist in individual document types that are derived from using different human languages, and that are implicit in the legislative culture of each country. Therefore the common open access model needs to be

designed to be flexible, support exceptions, and allow extensions to provide support for all peculiarities that can be found in the complete document set.

4.1.2 National and transnational standards

A standard for legislative documents can aim only at being shared within a particular national jurisdiction, or also at being shared by different national jurisdictions.

Sharing a legislative standard within a single jurisdiction means that the standard can (and should) be used by all normative authorities in that jurisdiction (not only Parliaments, but also regional assemblies, regulatory bodies, local authorities, etc.), when producing and disseminating the documents. Moreover, it means that such a standard will be used by other authorities, private publishers, citizens, when accessing or reusing such documents. Finally the standard will be used by software developers, in order to produce software meant to support the preparation, the distribution, the use, the reuse of legislative information.

Sharing a legislative standard within a single jurisdiction will contribute to leveraging modern information and communications technologies to radically improve the usability of Parliamentary information as a strategic resource both inside and outside Parliaments. In particular, improved access to documents regarding Parliamentary activities will enable citizens to hold Parliaments accountable, stimulate greater efficiency and enhance democracy. We can consider Parliaments as major producers of data and information that are vital for the democratic well being of a country and the lifeblood of political participation. The lack of a standardized way for Parliaments to classify and structure their data resources, information technology and business processes stands in the way of increased integration of information exchange and this in turn limits the efficiency and effectiveness of Parliamentary activities.

Sharing a single jurisdiction-wide standard for legislative information fosters greater cooperation between different institutions, administrations and Parliaments using common open standards that avoid vendor lock-in and allow for greater public access to information.

Further increased and additional advantages can be obtained when a standard is shared not only within a single jurisdiction, but also by different jurisdictions, in the framework of a transnational cooperation. Parliaments are now endeavoring to promote collaboration and cooperation with Parliaments of other countries. A transnational shared standard would represent a significant contribution to overcome the present state, where most interactions within and among Parliaments require numerous disparate transactions across multiple departments/Parliaments and there is very limited consolidation and aggregation across national Parliaments' boundaries.

Connecting Parliaments has many benefits: in addition to its value as a knowledge transfer mechanism, where one Parliament can learn from the other, it can also be a tremendous boost to Parliamentary positive imitation. By seeing what others are doing, a Parliamentarian can discover the possibility of doing the same in his or her Parliament.

The explosion of Internet-based systems have increased the possibilities and range of such dialogue but this can be achieved and exploited only if common standards related to the production, classification and sharing of Parliamentary and legislative electronic documents are agreed and used by the Parliaments themselves.

4.1.3 Interoperability

Although each Parliament has its unique characteristics, all Parliamentary democracies have a number of characteristics in common: Actors, Structures, Procedures, Acts and Information. A good standard defines common building blocks in a single model that can be applied to each (or at least most) Parliamentary documents.

A good standard defines a set of recommendations and guidelines for e-Parliament services in an international context. The framework will be an essential prerequisite for interlinking and web-enabling Parliaments. It will address information content and recommend technical policies and specifications for connecting Parliament information systems across countries. It needs to be based on open standards.

Country Parliaments should use the guidance provided to supplement their national e-Government Interoperability Frameworks with an international dimension and thus enable international interoperability of Parliaments. So a good standard is meant to supplement, rather than replace, national interoperability guidelines that may exist by adding the interoperability dimension. Any such initiative will enable open access by focusing on both *semantic* and *technical* interoperability.

- Semantic interoperability is concerned with ensuring that the precise meaning of exchanged information is understandable by any person or application receiving the data.
- Technical interoperability is aimed at ensuring that all applications, systems and interfaces for e-Parliamentary services are based on a shared core of technologies, languages and technical assumptions easing data interchange, data access and reuse of acquired competencies and tools. A good standard ensures technical interoperability by enforcing the use of open standards and open document formats, such as those based on the XML (eXtensible Markup Language) language, whose specifications are a worldwide standard and for which numerous tools and applications have been developed and are widely available.

By adopting such specifications, Parliamentary system designers (and more generally, the designers or information systems for any system meant to support production, access and reuse of normative information) can ensure interoperability between systems while at the same time enjoy the flexibility to select different hardware, and systems and application software to implement solutions.

4.1.4 Presentation, structure and semantics

There are three aspects to any Parliamentary document: 1) Presentation - how the information looks e.g. the colour of the text used in the document, the headings and other such formatting

issues; 2) Structure - how the information is organized; 3) Semantics - what the information represents or means.

Online publishing of Parliamentary documents has long been confined to presentation issues. Documents have been put on line trying to replicate as much as possible the layout and formatting of paper. The way a document looks is very important to the “human reader” but does not really provide much useful information to the computer to actually “read” a document as a knowledgeable human being could do. The development of descriptive markup meta-languages such as XML allows adding information to any document that would make both the structure and the semantic of a document “readable” by a computer. Computer do not have the kind of experience and knowledge that allow professional human being to be able to deduct structure and semantics from a document unless this document has been previously “marked up” to make it “machine readable”. More specifically:

- Semantic markup – semantically identifies parts of the document (e.g., headings, names, references, provisions. In this way the “meaning” of the different parts can then be “understood” by machines as well, in the sense that a machine will be able to distinguish such parts and consequently to process them accordingly.
- Structural markup – this refers to the categorization of different parts of a document based on their functionality e.g. In a Parliamentary document you may want to indicate that a certain section of the document is the Preamble, Question, Motions etc.

A good legislative standard may provide a way to move digital documents from the presentation era into a more semantic-oriented generation, in which digital Parliamentary documents are not just displayed online, but are “understood” by software applications (according to the ideas of the semantic web, see Section 2.2. Both the “meaning” and “structure” of every element in a Parliamentary document is available for all machines to access, thus providing the unprecedented opportunity to exploit the speed and accuracy of ICTs to manage access and distribute such documents.

4.2 Scope of a standard for legislative documents

Parliaments function through the medium of documents. Debate in Parliamentary chambers is recorded as documents. Legislation is passed through the voting process via a combination of documents; the proposed legislation itself, proposed amendments, committee working papers and so on. Given that the process is document-centric, the key enabler of streamlined Information Technology in Parliaments will be the use of open document formats for the principal types of documents. Such open document formats will allow easy exchange and aggregation of Parliamentary information, in addition to reducing the time required to make the information accessible via different electronic publishing media.

In the following section we will consider the main conventions which concur in providing a standard for legislation, namely, a structural convention (concerning ways to partition and organize of the documents), a metadata convention (concerning ways to add information to the

document), a naming convention (concerning ways to name documents and their parts and reference them),

4.2.1 A structural convention

A good standard needs to include a comprehensive set of XML-based Parliamentary Open Document Formats. Document standards make use of industry standard XML (eXtensible Markup Language) to define open document formats. Parliaments work with a number of distinct types of documents such as legislation, debate record, Parliamentary questions etc. Thus, comprehensive set of XML-based Parliamentary open document formats would cover all of the following categories:

- Primary Legislation – covering the lifecycle of a piece of legislation
- Parliamentary Debates
- Amendment lists
- Committee briefs
- Journals

The standard should define a distinct document type for each major type of document. The definition takes the form of human and machine-readable document models, one for each document type. All document types would share the same basic structures, provide support for metadata, addressing and references, and differentiate common structure and national peculiarities and extensions. All documents can be produced by the same set of tools (although specialized tools may provide more detailed and specific help in specific situations), need the same tools to be displayed or printed (although specialized tools can provide more sophisticated and individual presentations), can reference each other in an unambiguous and machine-processable way, and can be described by a common set of metadata that helps in indexing, analysing and storing all documents.

4.2.2 A metadata convention and a shared ontology

Metadata is structured information about a resource. Metadata records information about a document that does not actually belong to it, but is necessary to examine in order to deal with it (for instance, information about its publication, lifecycle, etc.). Metadata also enables a resource to be found by indicating what the resource is about and how it can be accessed. Furthermore, metadata facilitates the discovery and use of online resources by providing information that aids and increases the ease with which information can be located by search engines that index metadata. Metadata values are labeled and collected according to a common ontology, i.e. an organized description of the metadata values that describe the resources. A common ontology is fundamental to provide a way for managing, organizing and comparing metadata.

The Parliamentary Ontology is concerned particularly with records management and resource management, and covers the core set of elements that contain data needed for the effective management and retrieval of official Parliamentary information. The aim of the Parliamentary Ontology is to provide a universal container for all the information about a resource that is available to the owner of the resource, does not belong to the resource itself, and might be

needed for management or searching. The Parliamentary Ontology needs to be designed to be extensible so that those Parliaments with different, or more specific, metadata needs may add extra elements and qualifiers to meet their own requirements.

4.2.3 A naming convention

The Parliamentary Naming Convention and the reference mechanism are intended to enable a persistent, location-independent, resource identification mechanism. The adoption of a scheme based on this Naming Convention will allow the full automation of distributed hypertext. The reference mechanism, based on a shared naming convention, allows the automated generation of hypertext links and access to resources explicitly cited in legislative documents. This automation can cater for:

- the availability, at a certain time, of more than one resource corresponding to the document referred to;
- the possibility that references to resources not yet published on the Web are present.

Official documents, bills, laws and acts contain numerous references to other official documents, bills, laws and acts. The whole Parliamentary corpus of documents can be seen as a network, in which each document is a node linking, and linked by, several other nodes through natural language expressions. The adoption of a common naming convention and a reference mechanism to connect a distributed document corpus, like the one embodied by the Parliaments, will greatly enhance the accessibility and richness of cross references. It will enable comprehensive cross referencing and hyper linking, so vital to any Parliamentary corpus, from:

- debate record into legislation
- section of legislation to section of legislation in the same act
- section of legislation to section of legislation in another act of the same Parliament or of other international institutions.

4.3 The users of a standard for legislative documents

A good legislative standard aims at providing support for a large number of tasks and users spread throughout time, space and competencies. The types of potential users that might end up using or benefiting from such standard can be grouped in the following categories:

4.3.1 The legislator

The legislator is either a member of a Parliament, or a personal assistant of a member of the Parliament. He is currently drafting a new piece of legislation, due to be discussed and, maybe, approved in a future session of the Parliament. “The legislator” is not aware of the existence of legislative standards, XML, or any such technicality. He/she might, or might not, be aware of the existence of guidelines in the formal drafting of law, he/she does not know what XML is, and does not care. He/she wants to be able to retrieve bills and acts effectively, to be able to access explicit references to other laws made in a bill or act, etc. The legislators want to be able to access “point-in-time” consolidations of laws that provide a consolidation of the original act and the subsequent amendments up to a specific point in time, the legislator wants

easy and effective tools to find and retrieve bills and acts to carry out their duties more effectively.

4.3.2 The legal drafter

The legal drafter is a member of the office supporting the process of legal drafting. During the discussion phase in Parliament, “the legal drafter” receives all proposed text modifications to a bill being discussed, and generates any of a number of documents used by members of the Parliaments (such as summaries, synoptic views of amendments, etc.). When the proposed bill is finally approved, he/she creates the final version of the bill; either directly in XML or in a word processing file that is then translated into XML by some downstream process phase. “The legal drafter” is an expert in the matters of law, and has some computing experience, but he/she is neither a lawyer nor a computer scientist. He/she knows something about legislative standards and XML, but not much.

4.3.3 The toolmaker

The toolmaker works for an ICT company who has a contract for creating specialized software for a Parliament. The toolmaker decides to create a specialized editing tool by customizing a well-known Word Processor (such as MS Word or OpenOffice) and a conversion tools that creates valid documents recognizing formatting characteristics of the input texts. He has the goals of making the tools usable by the legal drafter and his/her colleagues, and at the same time compatible with the rules of the standard. Differently from the “legal drafter”, “the toolmaker” has full access to the standard’s documentation, and can talk to his users to understand together what each part of it really is relevant to their task and how to proceed.

4.3.4 The citizen

The citizen of a country where the system based on the good standard is being used might be a lawyer, a judge, a public employee, or just any ordinary citizen who needs fast and easy access to laws and legislation for his/her own purposes. “The citizen’s” main objective is searching for laws either through an explicit reference (e.g. section 36(2)(c)(ii) of Act 2-1999) or via a search interface (either textual or exploiting vocabularies and ontologies specified through the legislative metadata). “The citizen” doesn’t know what the standard is about, nor what XML is, and does not care. He/she wants his/her web browser to display the text of law searched, he/she wants all explicit references to other laws to be hypertext links, and a reasonable interface that lets his/her read the text on the screen and, when necessary, print it on paper.

4.3.5 The future toolmaker

“The future toolmaker” is 10 years old now. He is playing with his school friends, does not know anything about legislative standards and does not care. Yet, he/she is in this list because in fifteen years, when turning 25, he/she will be a professional computer programmer and will have to create new tools for the Parliament. The key difference between “the toolmaker” and “the future toolmaker” is that “the future toolmaker” may not have access to complete documentation, lost through time after internal evolutions of the standard as well as the workflow of the Parliamentary systems. He/she will only have sparse documentation of the actual requirements of the system he/she is working on. Furthermore, he/she will have to deal

with a fairly stratified situation where the basic ideas (on which “the toolmaker” has worked) have evolved, modified, expanded and changed emphasis. Furthermore, more often than not these changes have happened slowly and without documentation. The only certainty that “the future toolmaker” has to work on is more than 15 years of legislation available in XML format, whose documentation is introductory for certain, but far from complete and sufficient. Fortunately the early decisions have been to have the XML format be as self-explanatory as possible, so that “the future toolmaker” can, in principle, deduce all undocumented facts about the standard by simply examining a few relevant XML instances of the legislation and discovering there how it should work. In a sense, “the future toolmaker” is more a key user for our system than “the toolmaker“, and the possibility for “the future toolmaker” to deduce fundamental properties of the standard from the visual examination of XML documents will make us sure of long-term existence and usefulness of the e-legislative system itself.

4.4 The tools for the standard-based management of legislative documents

Just as many are the users (some of whom are not even aware of the fact they are using or relying on XML-enabled systems), many also are the tools that need to be created around the standard document model. Some of them are basic tools that are necessary for the system to work at all. Others are additional applications that will be created once the basic tasks have been catered for. A brief list of the main categories may help in explaining the breadth and variety of scope of such a standard, and the number of issues that need to be considered in the development of the data formats.

4.4.1 The editor

The editor is the fundamental tool for the generation of XML versions of legislation. Although not all drafting needs to be actually done on a specialized editor (much less an XML editor) in any real life scenario, there will be situations in which that will be possible and actually necessary.²³ The editor will be used in three different scenarios:

- As an interface to activate, control and verify the automatic conversion tool previously described. Through the editor “the legal drafter” will be able to verify the correctness of the conversion, change and add whatever the conversion engine has forgot or misidentified.
- As a tool to manually mark up a document provided in a different format. Depending on the sophistication of the conversion engine, this scenario will most probably blend naturally with the first one. Surely the editor will provide for functionalities to edit and add any kind of conformant markup, and will be able to check validity of the intermediate result.
- As an application for direct insertion of both text and markup, starting off an empty document: this will probably be the rarest scenario of use, as the drafting offices will most usually work off an existing document in some other format.

²³ As examples of editors for generating standard-based legislative documents, see Palmirani and Brighi 2003, Biagioli et al. 2006.

4.4.2 The converter

It is the converter, with the editor, the most fundamental tool for bootstrapping a standard-based management of legislative information. It will take some effort to convince “the legal drafter” to switch from his/her old faithful word processor and his/her manual system of handling amendments through a combination of glue and scissors, to use any kind of strange text editor. In the meantime, one of the most important tools will be the converter (for an example of a converter, see Palmirani et al. 2005).

The converter has the double purpose of converting into XML structures the documents that “the legal drafter” is still producing traditionally, and, most importantly of all, of converting into XML structures the legacy documents, the already approved bills and acts that form the current legislative situation of the country, and whose conversion to XML is needed for any hypertext web of references to work at all. Since legacy documents are, by definition, in any old format, and since “the legal drafter” is not interested in converting them into XML using an editor, “the toolmaker” will have to create an automatic mechanism for the task anyway.

The converter is based on the idea of semi-automatic conversion, i.e., it has automatic processes to determine as correctly as possible the actual interesting structures, and has a manual process to confirm (or, if there is an error, to edit) the inferences made by the automatic process. In fact, this application could be even one of the modules of the editor, and use the editor itself for corrections to the automatic inferences of the converter. Of course, the amount of human editing is inversely proportional to the sophistication of the converter, and in theory large quantities of documents could be processed automatically with little or no manual intervention.

The converter works by examining the typographical and textual regularities of the document, and inferring a structural or semantic role for each text fragment. For every fragment that has no deducible structural or semantic role, the presentation characteristics will be recorded instead and it will be left to the human user to infer the structural or semantic role (if any) needs to be associated to the fragment.

Experience with existing legislative systems shows that the basic structure of a legislative act (sections, subsections, clauses, preambles, conclusions, attachments, etc.) can be inferred automatically with great precision and few errors. The most important semantic elements, references and dates, can also be deduced automatically with great precision as long as the human-readable text used for them uses one of a limited number of acceptable forms. More complex structural elements (explicit modifications, specialized terms, people, etc.) might be difficult to catch automatically, but not impossible.

4.4.3 Name resolvers

In Internet-based framework, legislative documents will be stored on networked computers and will be accessible by specifying their addresses. Yet these addresses are extremely dependent on the specificities of the architecture that will be in vogue or appropriate for the

economic and technical context of the moment. It is extremely inappropriate, therefore, that any content or structure that is planned to last for more than a short period of time is given direct access to the physical address of the document in the form that will be eventually used for display.

As we have seen in Section 4.2.3, an essential component of a standard for legislative document is a Naming Convention, namely, a mechanism for creating identifiers of documents that can be used for accessing content and metadata regardless of storage options and architecture. However, a Name-convention needs not (and should not) be used directly for accessing these structures, since it should specify an architecture-independent URI address for all relevant structures of the concerned standard. Thus, a naming convention needs to be combined with a Name resolver. A Name resolver is a software tool that can, given an architecture-independent URI, identify the resource being sought and provide the current architecture-dependent address that needs to be used at any given time for actual access. Name resolvers are either indirect (in that they redirect the client application to the current address of the requested document) or direct (in that they immediately provide the requested document by generating the actual address and requesting the document as a proxy for the initial client application).

4.4.4 Validation tools

The “legal drafter”, as soon as editing is finished, needs validation tools to check that the document complies with all requirements:

- A content and structure validator that checks the correctness of the document instance with regard to the XML-schema and other rules applicable to it.
- A reference validator that checks whether all references contained in the document already belong to the document collection and are correctly referenced.
- A metadata validator that checks whether the metadata stored with the document are correct and complete.

4.4.5 Document-management and other post-editing tools

To take full advantage of the information encoded in standard-compliant texts, an advanced document management system is required, with search engines, hypertext functionalities, visualizations and versioning facilities. Such a system will provide the standard-compliant documents to a number of browsers and applications which will increase and get more sophisticated in time. Among such application we can mention the following:

- retrieval facilities, from general search engines, to legally specialized search agents;
- workflow managers, using the information inside the texts to direct the Parliamentary procedures concerning such texts;
- publishing system, providing the texts in different media and formats;
- knowledge-based systems, using the text for providing explanation and advanced metadata (concepts and rules) associated to it for performing inference.

Such a list is certainly not exhaustive, since, as it has always happened with the Internet, once certain information is made available, multiple unpredictable ways of using it will be discovered by developers and users.

5 Conclusions (Giovanni Sartor)

In this final section, we shall shortly summarize the main results of our inquiry on ICT and the management of legal documents.

In the information society, Parliaments need ICT to be able to properly discharge their function. Legislative information system must meet the challenges of our time, such as the necessity to cope with the increase in quantity and complexity of legislative texts and with their accelerated change, but also the needs to enhance the quantity and quality of information available to legislators, to make legislative processes more transparent and open to citizens' participation, and to enable experts and interested parties to contribute with their knowledge and their perspectives to the discussion of regulatory problems. Finally, as international interdependence grows, problems become more complex, but also common to different countries. On the one hand the same problems (e.g. regulation of ICT or biotechnologies) appear in different countries, and on the other hand there are problems transcending the borders of a single country (e.g. environmental protection). This requires sharing legislative information for comparing legislative solutions and identifying best practices but also to coordinate efforts for addressing global issues.

Though the functions of legislative information systems cannot be limited to the production and management of documents, a central feature of parliamentary information systems consists indeed in its ability to manage legislative documents. These documents are not limited to the legally binding normative acts adopted by Parliaments, but also include all preliminary materials produced during the legislative process.

The evolution of legislative information systems shows some emerging trends:

- a comprehensive digital management of legislative document, which concerns all documents produced during the legislative process;
- the integration of digital documents into the legislative process, where such digital documents result from, and contribute to, the management of the legislative workflow;
- the opening of the legislative information system to society, where documents produced in the legislative process are made available to public access, and documents resulting from public debate and interventions are input into the legislative process;
- a diversified provision of legal information, where legislative documents are seen as a common societal resource to be distributed and reused in diversified ways, by public and private, profit and non-profit actors;
- the move toward the semantic Web, namely, the enrichment of documents with machine-processable information, which can facilitate access to and use of legislative document in unprecedented ways.

We have argued that the way forward consists in standard-based open-access to legislative documents, namely, in establishing a shared open standard for legislative document, so that these documents can be enriched with machine-processable information usable by any

computer system and any standard-compliant software. We have also argued that this information—covering the structure of the document, but also its identification, its description, its links, its lifecycle and aspects of its content—should be represented using the resources provided by the semantic Web (such as the metalanguage XML), and it should concern all documents produced during the legislative process.

Once such a standard is defined, a set of compliant tools needs to be developed in order to assist the preparation and the revision of standard-compliant documents, the management of their workflow, and their subsequent distribution and reuse. Some of these tools need to be used inside the Parliament, but their development need not be performed in-house, since the use of a shared standard (common to multiple legislative assemblies, possibly of different countries) will support the creation of a wide-enough competitive market. Moreover, the openness of the standard enables other actors to reuse and redistribute (possibly with added information) standard-compliant documents, contributing to the knowledge of the law and to participation in the legislative process.

In the two appendixes completing the present report, we shall further investigate approaches to the management of legislative documents: first we shall review initiatives dealing with standard for legislative documents and then initiatives dealing with legislative semantics.

6 Appendix 1. Managing legislative documents: a review (Enrico Francesconi and Monica Palmirani)

This appendix will provide a review of some significant initiatives in the management of legislative documents, focusing on the projects that are based on the idea of an open-access standard.

6.1 European initiatives

In the last few years a number of important initiatives in the field of legislative document management have been developed. Different national initiatives have introduced standards for legal source description as “Metalex” (Section 6.1.1) and SDU BWB (Section 6.1.2) in the Netherlands, “LexDania” in Denmark (Section 6.1.3), the “NormeinRete” project in Italy (Section 6.1.4), “CHLexML” in Switzerland (Section 6.1.5), “eLaw” project in Austria (Section 6.1.6).

Such initiatives mainly promoted by national institutions, involving also research institutes and universities, have defined XML standards for promoting standards in legislation management, as well as schemes for legal document identification.

A current initiative to define a European standard for legislation is being discussed as well (Section 6.1.10).

6.1.1 Metalex

Metalex has been developed as part of the E-POWER project. This project was aimed at the use of ICT to support citizens and governments in dealing with an increasing number of regulations. European citizens and enterprises are more and more confronted with rules and regulations. This affects various aspects of their business. Regulations come from international, European, national and local authorities. Despite attempts at harmonization and de-regulation, the size and complexity of the potentially relevant body of ‘law’ continuously increases. This is a problem for administrations too, legislative and executive bodies alike. The process of drafting consistent and coherent legislation is getting more complicated, as is that of upholding and applying valid law. ICT has the potential of supporting both the government and citizens in dealing with this increasing body of law.

A precondition for the tools that were developed in E-POWER was the electronic availability of legal sources in a structured and standard format. Metalex was developed to fit this need Boer et al. 2002, Boer et al. 2003. In addition, it allows for exchange and comparison of legal documents from different sources (such as publishers). It provides a generic and easily extensible framework for the XML encoding of the structure and contents of written public decisions and public legal documents of a general and regulatory nature. Metalex poses only minimal requirements on the structure of documents.

Currently, Metslex is used by the Dutch Tax and Customs Administration, Be Value, the Belgian Public Centers for Welfare and others.

6.1.2 SDU BWB

Other institutional initiatives in The Netherlands is the Dutch SDU BWB standard. It is an XML format currently used for encoding the laws in the Dutch Basiswettenbestand (BWB) database, that is a large database containing almost all Dutch laws and decisions. The standard is based on a DTD originally developed by SDU publishers, and now maintained by the Dutch government.

6.1.3 LexDania

The LexDania project is defining a national Danish system for the creation and interchange of legislative documentation. LexDania was initiated by the Danish Ministry of Science, Technology and Innovation then continued by the Retsinformation (Ministry of Justice) and the Folketinget (Danish Parliament). The work was conducted in two phases. In a first phase, a research on international activities and an investigation in other national standards and projects was done. In the second phase, the development followed using the basic data model / methodology of the General Danish Public Information Online (OIO) XML strategy. This strategy consisted in choosing a set of central types and elements (standards if possible), creating sets of “building blocks” for national use, (re-)using building blocks to create specific legislative schemas Tucker 2004b, Tucker 2004d, Tucker 2004e, Tucker 2004c, Tucker, Tucker.

The project is focusing on developing a system of schemas for the systematic creation and maintenance of document type and application schemas. The system has a unique approach to building schemas. A structure of stratified layers is used to incrementally construct the schemas from functional features – rather than document characteristics. The structure is accompanied by a methodology explaining ways of constructing schemas to assure consistent and compatible schemas.

6.1.4 NormeinRete

The project NormeinRete (Lupo and Batini 2003), or NIR, started in 1999 with the leadership of CNIPA together with the Italian Ministry of Justice, and it gathers several Italian public institutions and research organizations.

During the last few years, NormeinRete has reached the following results:

- a web portal (www.normeinrete.it) providing a unique access point for searching the Italian legislative corpus. It offers search and retrieval services operating on all Italian laws since 1904, and utilities for automated hyperlinking. The entire project documentations and other information related to the project are also available (in Italian). The portal includes e-learning facilities, a software download section to deliver open source utilities developed by the project

team, and a best-practices section to encourage experiences re-use, in order to create a virtual space for knowledge sharing within the Public administrations community.

- a standard for XML representation of legal documents (Megale and Vitali 2001). DTDs and XML Schemas for Italian legislation have been defined. These schemas can represent metadata Biagioli 1997 and all the significant information useful to automate legislative documents life-cycle management. Moreover, the availability of XML documents marked-up according to shared formats allows to provide advanced search and retrieval functions operating on heterogeneous data bases effectively.
- A standard for persistent identification of legal documents, compliant with the IETF Uniform Resource Name Moats and Sollins 1997, and an infrastructure for identifiers resolution and management (Spinosa 2001).

The NormeinRete standards have been issued as Technical Norms by Italian Authority for information technology in the Public Administration (AIPA – now CNIPA), and published on the Italian Official Journal AIPA 2001; 2002.

Nowadays, both standards have been adopted by the majority of Italian public administration and by a growing number of private operators. For example, the whole legislative documents collection of Italian Supreme Court of Cassation, the most wide and prestigious national collection in this field, adopt NormeinRete standards.

NormeinRete is the result of several years of collaboration between technology and law scholars, software developers, public administrators both at the national and regional levels.

6.1.5 CHLexML

The activities on standardization in legal domain, with particular attention to multilingual issues in the Swiss Confederation, are coordinated by COPIUR “Coordination Office for the Electronic Publication of Legal Data Federal Office of Justice” in Bern, Switzerland²⁴ ().

Copiur 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 (Khaled et al. 2004), 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.

²⁴ <http://www.rechtsinformation.admin.ch/copiur/index.html>

COPIUR currently deals with two relevant projects:

The first one is CHLexML²⁵, 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 finalized 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 5th seminar on law and informatics organized by Copiur on 30 June – 1st 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 normalization.

The second project is named LexGo²⁶. 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.

6.1.6 eLaw

The Austrian institutional initiative eLaw Project aims at a reform of legal text production, creating a continuous electronic production channel with a uniform layout prepared on the same electronic text basis from draft to publication (promulgation) on the Internet. The workflow system includes government bills, committee reports, legal enactments of the Nationalrat and decisions of the Bundesrat. Parliament returns consolidated electronic texts of legal enactments to the Federal Chancellery, ready for publication. The authentic electronic publication on the Internet (since 2004) is available for everybody free of charge.

The eLaw (e-Recht “Electronic Law”) project aims at creating one continuous electronic production channel from the invitation to comment on draft legislation to promulgation (on the Internet). As a result, it is only required to enter amendments to the text during the legislative stages (for example by a committee, or in the plenary of the Nationalrat). As the first result of the project, texts of laws on paper are to be replaced by electronic texts, that is to say, printed government bills, committee reports and other parliamentary printed matter will cease to exist. Technology will make it possible to draw up texts, which can be queried electronically while all stages can be tracked in a fully transparent process. Primarily for the purpose of cost-cutting, the texts of legislation were to be given a uniform layout and were to be prepared on the same electronic text basis from draft to publication in the Federal Law Gazette on the Internet. As a result, the Federal Chancellery sends government bills to

²⁵ www.chlexml.ch

²⁶ www.lexgo.ch

Parliament; Parliament returns the consolidated electronic version of the legal enactment adopted by the Nationalrat once parliamentary procedures have been completed. The State Printing Office (Wiener Zeitung) is no longer necessary.

To translate eLaw into reality, two projects were launched by the Administration of Parliament in view of the complex task on hand and the brief period available by decision of the federal government (trial operations were to start as early as on 1 September 2001):

- the “Implementing E-Law” project (in April 2001) to ensure one continuous electronic channel for the legislative procedure in the Nationalrat and the Bundesrat, as well as
- the “Roll-out Plan for laptops’ to be used by Members of Parliament” (in December 2002).

The basic ideas of the eLaw project are:

- To provide an electronic workflow for producing legal texts beginning with the draft bill and ending with the ePublication of the Federal Law Gazette (e.g. law, regulation, announcement, treaty)
- To replace printed legal texts by digitally signed electronic documents
- To officially publish the Austrian Federal Law Gazette on the Internet

The exchange of documents with the federal administration based on compatible formats and on independent interface, and detailed procedures have been developed for the exchange of data between parliament and government.

The system offers the people involved in the legislative process not only electronic information but also participation opportunities. In this context, it has to be mentioned that a special upload mask was created for Members of Parliament and their assistants to enter the electronic versions of motions into the system. In a competence center established in Parliament, which has meanwhile successfully completed know-how transfer from the State Printing Office and taken up cooperation with the Federal Chancellery, staff supports the rapporteurs of the committees and the committee secretaries of the Parliamentary Administration in preparing the committee reports as well as the staff members of the Parliamentary Administration responsible for executing the legal enactments of the Nationalrat, and is in charge of quality management and the layout of legislative documents as well as covering the need for additional labour in peak times. The electronic exchange of data between government and parliament takes concrete shape as follows: the Federal Chancellery sends government bills to Parliament, Parliament returns the consolidated electronic version of any legal enactment adopted by the Nationalrat once parliamentary procedures have been completed.

6.1.7 Legal XML

The Legal XML mission is to develop open, non proprietary standards for legal documents and associated applications. LegalXML is a collection of standards developed by different Technical Committees, covering a wide spectrum of legal materials. To date, the Court Filing Workgroup is the first and only workgroup to publish a “proposed” standard specification.

6.1.7.1 Context from which XML language started

Formed in 1998, LegalXML is a collaboration of attorneys, court administrators and IT staff, academics, and companies. Early LegalXML work focused on electronic filing of court documents. The group's first specifications, "*Court Filing 1.0 and 1.1*", "*Court Document 1.1*", and "*Query and Response 1.0*" addressed key areas of concern for attorneys and managers of court case records.

LegalXML joined OASIS (the Organization for Advancement of Structured Information Systems) in 2002. Much of the work done in LegalXML touches on the concerns of national organizations and associations outside OASIS. Such work is presented to those groups for consideration for their adoption as business and technical standards. Specifications from the Electronic Court Filing Technical Committee (ECFTC) are submitted for review and have been adopted by the Joint Technology Committee (JTC) of the Conference of State Court Administrators (COSCA) and National Association for Court Management (NACM).

6.1.8 LAMS

The "Legal and Advice Sectors Metadata Scheme" (LAMS²⁷) initiative, developed by the Lord Chancellor's Department in UK as part of the Community Legal Service (CLS) launched in April 2000, aims at promoting common standards across Internet sites developed by organizations in the Legal and Advice Sectors.

It is specifically concerned with the standardization of websites holding information on legal matters to the extent that they should classify information according to a common framework. The proposals are intended to deliver benefits to both providers and users of these websites.

LAMS conforms to the standard metadata element set of "Simple" Dublin Core. This gives the greatest possibility of gaining the advantages offered by the adoption of an existing metadata scheme.

6.1.9 UKMF

UK Metadata Framework (UKMF) aims at describing 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.

6.1.10 The Metalex/CEN initiative

On the basis of such experiences, in 2006 an initiative has been launched at CEN (European Committee for Standardization) for a Workshop on Open XML interchange format for legal and legislative resources. This initiative will be discussing also the definition of a unique identifier for legal measures.

²⁷ www.lcd.gov.uk/consult/meta/metafr.htm

A CEN Workshop Agreement (CWA) on an Open XML interchange format for legal and legislative resources has been accepted by the CEN and associated standard organizations as a publicly available specification for the period of three years, after which the agreement must be renewed or upgraded to a norm.

The MetaLex standard is considered as an interchange format between other, more jurisdiction-specific XML standards. As such, it is very abstract. Therefore, it is considered a basis to develop the new standard called MetaLex/CEN schema. It is based on previous versions of the MetaLex schema, the Akoma Ntoso schema, and the Norme in Rete schema. Other relevant parties are i.a. LexDania, CHLexML, FORMEX, etc. In addition to these government or open standards there are many XML languages for publishing legislation in use by publishers.

6.2 African initiatives

6.2.1 AKOMA NTOSO

In 2004 and 2005, the United Nations Department for Economic and Social Affairs (UN/DESA) project “Strengthening Parliaments’ Information Systems in Africa” aimed at empowering legislatures to better fulfill their democratic functions, using ICT to increase the quality of parliamentary services, facilitate the work of parliamentarians and create new ways to promote the access of civil society to parliamentary processes. In the first stage, the project has been actively supported by the NormeinRete community and has been influenced by NormeinRete project results.

A strategic role in this project is played by the AKOMA NTOSO²⁸ (Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontology) framework, a set of guidelines for e-Parliament services in a Pan-African context. The framework addresses information content and recommends technical policies and specifications for building and connecting Parliament information systems across Africa. In particular, the AKOMA NTOSO framework proposes an XML document schema providing sophisticated description possibilities for several Parliamentary document types (including bills, acts and parliamentary records, etc.), therefore fostering easier implementation of Parliamentary Information systems and interoperability across African Parliaments, ultimately allowing open access to Parliamentary information.

AKOMA NTOSO was developed as a necessary foundation for the development of a comprehensive Parliamentary Information System (PIS). The goal of the Parliamentary Information System is to maximize the operational efficiency and effectiveness of National Legislatures by implementing a solution which provides secure, reliable, and timely collection, storage, access, and transmission of information. The aim is to equip Parliaments

²⁸The “Akoma Ntoso” (“linked hearts” in English) symbol is used by the Akan people of West Africa to represent understanding and agreement

with a solution that fosters accessibility, transparency and accountability of Parliaments by exploiting open source multi-platform applications based on open standards and available in multiple human languages.

Individual country Parliaments are meant to use the guidance provided by the AKOMA NTOSO Framework to supplement their national e-Government initiatives with a Pan-African dimension and thus enable Pan-African interoperability of Parliaments. Thus the AKOMA NTOSO framework is meant to supplement, rather than replace, national interoperability guidance that may exist, and to add a pan-African dimension to them.

The AKOMA NTOSO Framework reaches three main objectives:

- to define a common standard for data interchange between parliaments;
- to define the specifications for a base document model on which parliamentary systems can be built;
- to define an easy mechanism for citation and cross referencing of data between parliaments.

The AKOMA NTOSO framework aims at providing two basic types of interoperability: semantic interoperability, which is concerned with ensuring that the precise meaning of exchanged information is understandable by any person or application receiving the data; technical interoperability, which is aimed at ensuring that all AKOMA NTOSO-related applications, systems, interfaces are based on a shared core of technologies, languages and technical assumptions easing data interchange, data access and reuse of acquired competencies and tools. AKOMA NTOSO ensures technical interoperability by enforcing the use of open standards and open document formats.

6.3 Asian initiatives

With regard to XML standards for legislation, Asia is characterized by initiatives carried on by University departments.

For example, recently, the Graduate School of Information Science of Nagoya University in Japan has carried out a study on legislative consolidation on a statute database (see Ogawa et al. 2007).

In Japan, some electronic databases of statutes have been established²⁹; however, they usually include only the current versions at the time when retrieved, when the database was established, or when the statutes in it were enforced. Some databases include former versions; more often they contain the version of the statutes when the databases have been established. Therefore to store every version of statutes in a database, a large number of documents has to be restored in a digitalized form.

²⁹ e.g., at the House of Representatives, see http://www.shugiin.go.jp/index.nsf/html/index_housei.htm or at the Ministry of Internal Affairs and Communications, see <http://law.e-gov.go.jp/cgi-bin/idxsearch.cgi>

This is not an easy task for the large amount of statutes is large and increasing. In fact, for example, among the acts enacted by the Diet, about 1800 are currently effective. In addition, more than 12000 acts have been enacted during the last 120 years, and about 200 new acts including about 150 amendment acts are enacted every year. In addition to these acts, about 5500 orders and regulations enacted by the cabinet and ministries are currently in effect.

Another problem exists. In Japan, as in many countries, amending a statute is performed by enacting an amendment statute, where the details of the amendment are clearly described in terms of amending sentences. Then the amending statute is consolidated with the current version to generate a new one. Thus, in principle, by repeating consolidation from the first version to the current one, every version can be acquired. A new version, however, sometimes cannot be easily and correctly obtained since this task needs knowledge about consolidation. In fact, in Japan, consolidation has so far been achieved manually by experts on legislation as paper-based work, and the knowledge must usually be acquired from technical guidebooks on legislation (e.g., Maeda 2003) or from other experts.

These problems suggest the necessity of an automatic consolidation system for statutes. Therefore, the Nagoya University has proposed an automatic consolidation system for Japanese statutes based on the formalization of amendment clauses, which are parts of amendment sentences, which can be formalized in terms of sixteen kinds of regular expressions.

The system utilizes XML techniques. Generally, a Japanese statute has logical structure elements such as a title, contents, the main provision, supplementary provisions, etc. The main provision has also a hierarchical structure that consists of parts, chapters, sections, sub-sections, and divisions. In addition, the elements below chapters are also hierarchical, whose substructure includes articles, paragraphs, items, sub items, etc.

For this purpose, a DTD has been designed to specify these structure elements of Japanese statutes and marked-up the statute documents based on this DTD. Each amending action is implemented as an operation to string in the texts or to the document structure.

Amending actions are classified into the following ten kinds:

1. Actions on strings in a statute text: (a) Replacement, (b) Addition, and (c) Deletion
2. Actions on structure elements of a statute such as sections, articles, items, etc.: (a) Replacement, (b) Addition, and (c) Deletion
3. Actions on numbers of structure elements: (a) Renumber, (b) Attachment, and (c) Shift
4. Simultaneous execution of several actions in the above, especially replacement of title strings and renumbering structure elements.

They are codified as regular expression and the automatic consolidation system implements such action to produce the consolidated acts.

Experiments have been executed on statutes to be amended, consisting of the first versions of seventeen acts enacted since 1947. With each of them, at least one and at most sixteen amendment acts need to be consolidated to obtain the current versions. From these amendment acts, 965 amendment clauses were extracted.

Then the final version of each act generated by the system as compared with the current version obtained from the existing database. The compared texts are leafs when the structure of the acts are regarded as trees. Among 4355 texts, 4332 resulted identical, proving the validity of the approach.

6.4 US initiatives

US is characterized by both institutional and non-institutional initiatives aiming at promoting legal standards, in particular some important private initiatives are currently acting to promote standards for semantic Web in legal communities.

6.4.1 An institutional case: XML at the US House of Representatives

The legislative drafting service (HOLC) of the US House of Representatives is provided by the Office of the Legislative Counsel. The Office provides drafting and related assistance to the Members of the House, the House committees, and the conference committees between the House and the Senate. The signed paper version submitted to the Clerk of the House on the House Floor is the official document of record.

HOLC consists of approximately 35 attorneys and a support staff of about 15 individuals, and is headed by the Legislative Counsel of the House who is appointed by the Speaker of the House.

Because the paper version is the document of record, the drafters provide their clients with typeset drafts or PDF files that can be printed in the client's office. The paper version of legislation is currently created in one of two ways:

- using a DOS editor, Xywrite, the drafter creates a file that contains typesetting codes along with the legislative language. This file is then processed through a typesetting program developed by the Government Printing Office (GPO). The output of the typesetting program is an Adobe PostScript file that is subsequently printed or distilled into a PDF file;
- Alternatively, using Corel's XMetaL application, the drafter creates an XML file that is converted to the typesetting coded file and processed as described above.

The transition to XML for the drafting of legislation has been both challenging and highly rewarding for the House. At this point, the House has been using the XML authoring environment for House-only resolutions since January 2001 and began drafting bills in XML

in September 2002. The House plans to draft over 95% of introduced bills in XML by January 2004. Through the Library of Congress' Thomas website³⁰ and GPO's GPO Access website³¹, the public accesses legislative documents. The Library of Congress is working on improved document display technologies using the XML files and GPO has developed a new version of Microcomp that typesets the XML files as well as locator coded files. DTDs, schemas, and examples of XML bills and resolutions are available at <http://xml.house.gov/>. In particular DTD and XMLSchema versions of XML standards for bills, resolutions and amendments have been defined.

The development of an XML authoring environment for drafting legislation has provided the opportunity to create a “smart” editing environment based on the context provided by the underlying XML structure. In addition, small changes in the drafting approach have provided opportunities for improved efficiencies. The goal for the new environment has been:

- Minimize drafters' attention to the typesetting product,
- Maximize drafters' consideration of the legislative language itself,
- Provide “just-in-time” knowledge support during the drafting process, and
- Provide all this within a WYSIWYG (non-Tags-On) environment.

6.4.2 A non-institutional case: Legal-RDF

One of the most significant non-institutional initiatives in the field of legal standardization is “Legal-RDF” McClure.

Legal RDF is a non-profit organization that is sponsored by legal firms, software companies, and other stakeholders interested in software tools that leverage the Semantic Web.

Advancing in this direction, the legal community will lower its costs; improve the quality of its services; and create an environment conducive to mass-customization of legal products.

The Semantic Web is a disruptive technology in this sense for the legal industry. Firms can grow significantly by fielding products that cater to the needs of clients who ordinarily would not pursue legal advice. Reasoning-based software offers opportunities to provide these clients a level of service with an acceptable level of risk.

Therefore the strategy of the Legal-RDF community is to construct two databases – a comprehensive open-source ontology that is then applied in structured descriptions of statutory and administrative codes. These databases are then leveraged by the Semantic Web community to create the reasoning software envisioned for orderly industry growth.

Specifically underlying this strategy is the cognition that functional requirements that apply to many legal documents (e.g., to contracts and wills) are nearly identical to those in legal

³⁰ <http://thomas.loc.gov/>

³¹ <http://www.access.gpo.gov>

statutes. Benefits will powerfully ripple through the entire community if relevant software is applicable equally to both domains. Consequently, development of a contracts-related ontology, separate from or preceding a statutes-related ontology, is undesirable from economic, social, and legal perspectives.

6.5 Australian initiatives

6.5.1 EnAct

Australia has introduced ICT for legislative processes since the second part of the '90s with the EnAct project. It involves the Tasmanian government, the Canadian government, some federal states of the US and New Zealand government. EnAct is a legislation drafting, management and delivery system that has been built to enable the Tasmanian Government to provide improved legislation information services to the community (Arnold-Moore et al. 2002). En-Act provides the community with a facility that enables cost effective public access to reliable, up-to-date, searchable consolidated Tasmanian legislation. The “point- in-time” capability allows users to search and browse the consolidated database as it was at any time since 1 February 1997. Tasmania achieved these goals by automating much of the legislative drafting and consolidation process.

The EnAct repository makes use of the Structured Information Manager (SIM) to store SGML fragments and associated metadata directly. Acts are stored in SIM as fragments with a timestamp marking the time interval (start and end time) over which the fragment or table of contents is valid. The SIM repository incorporates an SGML parser allowing sophisticated indexing based on the logical structure of the SGML fragments. Time point searching is also supported. When a time point is specified, a filter is applied to the database leaving only those fragments and tables of content that were valid at the specified time. This allows a snapshot of the database to be searched and browsed as if it were the entire database. The SIM web server uses the same SGML parser. This allows HTML to be generated dynamically from the fragment repository for delivering Web pages. This means that the same repository can be viewed in a number of different ways. In the current public site one can view a fragment with or without history notes. A single fragment can be viewed on one side with a table of contents on the other, or with the table of contents and all of the fragments united together into a single document. All cross-references are activated as hypertext links. Because all hypertext links are activated using queries in the database, it is just as easy to do reverse hypertext links—i.e. show a list of all fragments that refer to this fragment or this Act. When history notes are displayed, those that refer to amending Acts on the system are also hypertext links to those Acts. The table of contents can be viewed either as a conventional section number and head note or with a section number and list of links to successive versions of the corresponding fragments. This provides a quick overview of the history of a provision. A new feature is a previous and next version button for each fragment.

In addition to the public website, which delivers HTML to user's browsers, the Printing Authority of Tasmania (PAT) also uses the same repository to generate “on-demand”

authorized reprints. A member of the public can contact the PAT shop and request a reprint of a particular Act on a specified day. They search the repository to find the relevant table of contents and fragments at that time point, which are joined together into a single SGML document that is translated into Rich Text Format using SIM formatters. This RTF representation is then passed to high-speed printers for paper output. Each authorized version is preceded by a certificate of authenticity from the Chief Parliamentary Counsel.

At the end of the document two tables appear. The first shows the name and the time of commencement of all amending Acts applied to that Act since it was enacted. The second lists each provision that has been amended and how it was amended. The database that is shared by the public Web gateway and the private Printing Authority gateway is not the only repository in the EnAct system. A working database is also kept in the OPC. This database contains everything in the public repository, but also contains the politically sensitive draft Bills in preparation, and all of the workflow information about the status of Bills and other draft legislation. Data is periodically migrated from the production database to the publication database providing a level of protection from intrusion and preserving the integrity of the data on the public repository.

In the past, any consolidation has been a laborious manual process. Drafters or clerical staff were required to handle the text of an Amending Act and go through it, section by section, applying each amendment to one or more Principal Acts. In the past drafters composed amending legislation directly, combining the process of deriving the effect of the amendments with the text describing them. The Tasmanian government presented a different approach. The drafters would mark amendments directly on a consolidation of the Principal using strike-through and underline markings familiar to many lawyers. Amendment wordings for those markings would then be generated automatically.

These markings are then captured in an internal (SGML) representation of the changes called a Change Description Document (CDD). These changes are then used to generate amendment wordings, which are appended to a stub or substantive Bill Arnold-Moore 1997. This process is managed by a workflow enactment service that keeps the CDD and generated amendments together so that, if and when the amendment Act commences, the amendments can then be automatically applied to the principle to generate new fragments in the historical repository.

6.5.2 JSMS

Justice Sector Metadata Scheme (JSMS) is a development of the AGLS metadata scheme, and it is designed for the use of organizations in New South Wales, Australia, which are publishing legal materials on the Internet. JSMS makes some minor qualifications to the “Simple” Dublin Core metadata elements adopted by AGLS. It does not use any of the AGLS-specific elements but does add several of its own.

7 Appendix 2. Managing of legislative semantics: a review (Mariangela Biasiotti and Enrico Francesconi)

This part of the report will be devoted to the analysis of the state of the art of the linguistic resources used by the information technology, as instruments for overcoming the problems of access and knowledge of the legal information. These instruments developed for different linguistic and legal systems are also a methodological necessity to approach the ever growing problems related to multilingualism in legal text, to the harmonization between EU and National legislation and to the comparative analysis of Law.³²

Generally speaking within the definition “linguistic resource” three kind of strategic tools are to be identified among those existing: thesauri, semantic nets (or taxonomy) and ontologies.

More specifically and in brief, ontologies are controlled vocabulary expressed in an ontology representation language (OWL), whereas Taxonomies or Semantic nets are collection of controlled vocabulary terms organized into a hierarchical structure and Thesauri are networked collection of controlled vocabulary terms. Each term in a taxonomy is in one or more parent-child relationships to other terms in the taxonomy while in a Thesaurus associative relationships are used in addition to parent child relationships. Generally speaking, one has to bear in mind that at European level, apart from EUROVOC semantic standard, there are no common and shared semantic standards and that there only are some good practices at national level in almost each member state, facilitating citizens to access legislative information.

7.1 State of the art of emerging technologies

In our IT-based society, knowledge and information is spread out through a vast number of documents featuring the most different languages, shapes and typologies. The weak structure of these documents’ contents requires two basic steps to guarantee an effective management of information. Firstly, documents selected are those important to the users’ needs; secondly, information is extracted from the texts. These processes are especially difficult in the legal field, as law is characterized and conditioned by cultural and structural differences, as well as by the complexity of the system.

In many countries, public institutions, as the main producers and distributors of legal source of information, have promoted projects aimed at improving the availability and the free access to legal information via the Web as a significant component of the process of transparency in citizen and institution interaction. In this environment, information technology, in particular Natural Language Processing (NLP), allows access to digital contents.

³²This section is mainly based on F. Bargellini, M.T. Sagri – Deliverable D2.1 “Analysis of legislative processes and linguistic resources for drafting” of the DALOS Project, revised by D. Tiscornia and L. Serrotti.

The problem linked to acquiring and managing knowledge depends on the fact that documents diffuse their several creative features only through language, and using NLP processes, computers acquire an advanced ability to process language to decode messages. With the NLP techniques, it is possible to build, almost automatically, representations of the documents' contents, to support cross lingual information retrieval, information extraction, text mining and concept research.

On the basis of this access process there are three main technology typologies.

The first group is made of patterns and standards to structure texts, which, through XML technologies, allow the identification and description of all the structural components of legislative texts, thus supporting the definition of a standard meta-model and a deeper integration among the modules processing the language, steps that are needed to ensure interchangeability of information.

Within the LegalXML Community, it has been reached a common agreement on the basic elements of a Standard XML Framework for legal documents, and tools implementing such standards at national level have been developed. In Italy the *XmLegesEditor*³³ is a legislative drafting tool developed, owned/provided and maintained by CNR-ITTIG, and in the Netherlands, *MetaLex* general purpose editor (*MetaVex*)³⁴ is owned/provided and maintained by the Leibniz Center for Law of the University of Amsterdam.

The second group is made up by tools for the linguistic analysis of the texts, and the dynamic acquirement of knowledge, such as morphological analysers and parsers for the automatic acquisition of terminology and semantic information from texts. Among the most remarkable examples of such tools we wish to mention the *SALEM*³⁵ (*Semantic Annotation for Legal Management*), a NLP-based system developed within the *Norme in Rete* (NIR) project³⁶, T2K

³³Agnoloni T., Bacci L., Francesconi E., Giardiello G., Spinosa P., Uccheddu F., *XmLegesEditor v.2.0 rc5. Editore per la redazione di testi normativi*, Ittig-Cnr, Firenze 2006

³⁴Saskia van de Ven, Rinke Hoekstra, and Radboud Winkels. *MetaVex: Regulation drafting meets the semantic web*. In Michel Klein, Paulo Quresma, and N ria Casellas, editors, Proceedings of the Workshop on Semantic Web technology for Law (SW4Law 2007), June 2007.

³⁵Bartolini R., Lenci A., Montemagni S., Pirrelli, V., Soria C.: *Semantic Mark-up of Italian Legal Texts through NLP-based Technique*. Proceedings of LREC 2004, Lisboa, Portugal, 2004.

³⁶NIR is a project promoted the National Center for IT in Public Administration (CNIPA), launched in 1999 as part of the Italian E-government Plan.

System design, by now consolidated, consists of classes of XML DTDs for structuring normative texts and of metadata., the most relevant part of which deals with the formal/structural features of each type of source, and with urns for the identification of the partitions of texts. A detailed description of DTDs for legislative texts is published at: http://www.normeinrete.it/standard/standard_xml.htm;

(*Text-2-Knowledge*) implemented in the *TRAGUARDI* project of the department of Public Functions FORMEZ³⁷ and the GATE environment.

SALEM is a module for the automatic annotation of the logic structure of law-making documents, integrated in the XML-Leges editor, developed by CNR's 'Istituto di Teoria e Tecnica dell'Informazione Giuridica'. Through the computer analysis of the text, SALEM manifests the most relevant aspects of the regulations, spotting elements such as the addressee of the regulation, the sanction provided for, and the obligations due. These content elements are clearly noted in the text with the XML metadata, thus offering an easier management and information extraction.

The third group is made up by linguistic resources, that is, thesauri, computational lexicon, multilingual semantic nets (or lightweight ontology), and ontology-based technologies. When describing the state of the art in terms of European projects created with the aim of facilitating access to regulatory information, it is opportune to focus on the necessity of integrating structural documentary standards with semantic ones for the description of content.

7.1.1 Overview of language and knowledge resources: Thesauri, lightweight and foundational ontologies

The access to the legal information can greatly profit from the support of linguistic resources providing sets of language data and descriptions in machine readable form, such as written or spoken corpora and lexical resources (annotated or not), multimodal resources, grammars, terminology or domain specific databases and dictionaries, ontologies, multimedia databases, etc. So the majority of public actors choose to offer to the user (citizens) facilities to access the legal information, adding to the digital documentation semantic resource. Linguistic resources can be classified according to three different typology: thesauri, semantic nets and ontologies.

³⁷The TRAGUARDI project, directed by Anna Gammaldi of FORMEZ, supports public administrations to manage structural funds. Calzolari N., Lenci A., *Linguistica Computazionale, strumenti e risorse per il trattamento automatico della lingua*, in Riv. Il mondo digitale, n. 2, 2004.

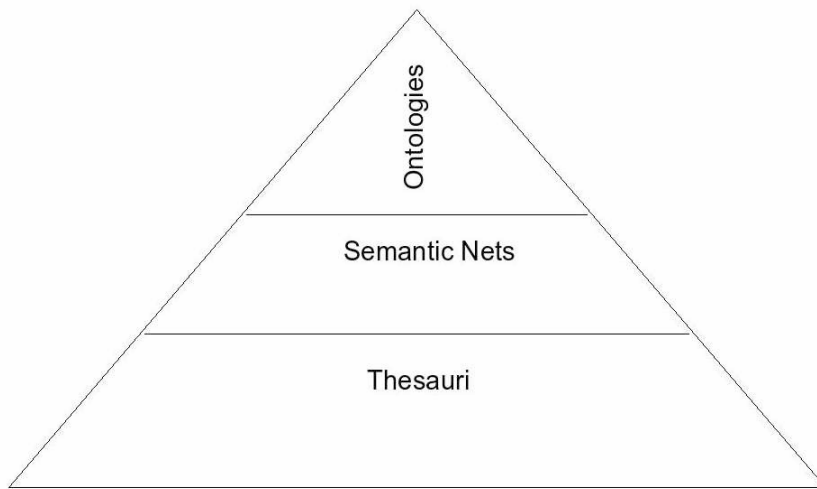


Figure 13: Use of Language Resources Pyramid

At a basic level of the Language Resources pyramid we find the Thesauri which, in general terms, are at the very first level due to the fact that they are the most widespread and utilized language tool. They can be defined as a *classification tool* to assist libraries, archives or other centre of documentation to manage their records and other information. This tool is designed to facilitate users to identify preferred (or authorized) terms for classifying and titling records and to provide a range of paths to reach these terms. The thesaurus also facilitates strategies for retrieving documents and reduces the probability of an unsuccessful research, or one resulting in a confusing or irrelevant outcome. This functionality is achieved by establishing paths between terms. The establishment and development of a thesaurus is generally arranged in accordance with the standards of ISO (International Standards Organization), which are officially recognized at international level. The definition of thesaurus supplied by ISO guidelines is the following: a thesaurus is *the vocabulary of a controlled indexing language, formally organized so that the a-priori relationships between concepts (for example as “broader” and “narrower”)* are made explicit. In addition to this explanation we find another definition, related to the multilingual thesaurus, which is: *a thesaurus containing terms selected from more than one natural language. It displays not only the interrelationships between terms, but also equivalent terms in each of the languages covered.*

Going up in the pyramid of the semantic resource we find more advanced tool such as “lightweight ontologies” or semantic nets (such as *WordNet*, *FrameNet* and the *CYC*, described in the next paragraph) and highly formalized ontologies, also called foundational ontologies (such as SUMO³⁸ and DOLCE³⁹).

³⁸The Suggested Upper Merged Ontology (SUMO) and its domain ontologies form the largest formal public ontology in existence today. The Standard Upper Ontology (SUO) is an effort by IEEE members to create a high-level ontology, for use by expert systems and intelligent agents within a variety of domains. The goal is to create a framework by which disparate systems may utilize a

Ontology is a central concept in emerging ICT development and the ontology-based approach is one of the most promising solutions to the “semantic problem” in the Web.

The term “ontology” is a concept coming from philosophy, but in the IT context, in particular in the AI, it represents a research stream studying how to represent reality correctly to allow deeper sharing of information and knowledge. The term has acquired several specifications in the information technology context, such as *lightweight* and *formal, axiomatic ontologies*.

Lightweight ontologies or Semantic nets appear as simple taxonomic structures of primitive or composite terms together with associated definitions. Lightweight ontologies are often a kind of computational lexicons, and are used to represent semantic relationships among terms in order to facilitate content-based access to the (Web) data produced by a given community. In this case, the intended meaning of primitive terms is more or less known in advance by the members of such a community.

Lightweight ontologies, such as lexicons, show only a limited kind of formal modeling, as they are generic and based on a weak abstraction model; the elements (classes, properties, and individuals) of the ontology depend primarily on the acceptance of existing lexical entries. In a lexical ontology, such as WordNet⁴⁰, many of the hyper/hyponymy links are not logically consistent, as it was designed as a lexical resource⁴¹. In lexical ontologies constraints over relations and consistency are ruled by the grammatical distinctions of language (language-based approach).

However, the need to precise agreements as to the meaning of terms becomes crucial as soon as a community of users evolves, or multicultural and multilingual communities need to exchange data and services^{39,42}

common knowledge base from which more domain-specific ontologies may be derived. It theorized that having a generalized means by which to express concepts and axioms, will encourage interoperability, reuse and more robust ontology import/export [Kent 2003]. They are being used for research and applications in search, linguistics and reasoning. SUMO is the only formal ontology that has been mapped to all of the WordNet (<http://www.cogsci.princeton.edu/~wn/>) lexicon. SUMO is written in the SUO-KIF language. This put the ideals behind KIF into practice, using a standard interchange language to express different ontologies uniformly [Sowa 2000]. The SUMO is intended to express the most basic and universal concepts that exist, to ensure the greatest level of generality and applicability to various domains.

³⁹DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) is a foundational ontology (FO) developed originally in the EU WonderWeb project. FOs are domain-independent axiomatic theories, contain a rich axiomatization of their vocabulary, and are used in order to make the rationales and alternatives underlying different ontological choices as explicit as possible.

⁴⁰Fellbaum C., *WordNet: An electronic lexical database*, MIT Press, 1998.

⁴¹See (Gangemi et al., 2001), for an interesting analysis of the logical consistency of the top level hierarchies of WordNet.

⁴²Gangemi A., Battaglia, M., Catenacci, C. *The Inflammation Ontology Design Pattern*, in Pisanelli (ed.), *Biomedical Ontologies*, IOS Press, 2004.

To capture (or at least approximate) such subtle distinction we need an explicit representation of the so-called *ontological commitment* about the meaning of terms, in order to remove terminological and conceptual ambiguities. This is the role of *foundational ontology*.

Foundational ontologies are axiomatic theories about domain-independent *top level* categories, such as *object, attribute, event, parthood, dependence and space-time connection*. They amount to repositories of highly general information modeling concepts that can be reused in the design of application ontologies for all kinds of domains⁴³. By providing toolboxes of standardized knowledge representation primitives, foundational ontologies also enhance the semantic interoperability between the communicating agents.

According to Artificial Intelligence wisdom, ontologies is nothing other than “the formal statement of a shared conceptualization”⁴⁴ and *foundational ontologies* are the top-level reference ontologies, to drive the construction of *domain ontologies*; though their starting point is the set of common-sense intuitions that make up the human conceptualization of reality, they ultimately aim at describing the categorical structure of the world as a whole. Generally, in information science, an upper ontologies (top-level ontologies or foundation ontologies) is an attempt to create an ontologies which describes very general concepts that are the same across all domains. The aim is to have a large number of domain ontologies accessible under this upper ontologies. It is usually a hierarchy of entities and associated rules (both theorems and regulation) that attempts to describe those general entities that do not belong to a specific problem domain.

The role and nature of foundational ontologies (and axiomatic ontologies in general) is complementary to that of lightweight ontologies: the latter can be built semi-automatically, e.g. by exploiting machine learning techniques⁴⁵; the former requires more painful human labour, which can gain immense benefit from the results and methodologies of disciplines such as philosophy, linguistics, and cognitive science.⁴⁶

Generally in the concrete applications of ontologies, different views provided as well as the representation of the reality given by them can be combined and integrated by means of the *core ontology*, to a *domain ontology*. To give an example, the ontology of a specific regulatory field (consumer protection, IPR Management, private law) is indirectly linked to

⁴³Gangemi A., Guarino N., Masolo C., Oltramari, A., Schneider L., *Sweetening Ontologies with DOLCE*, in Proceedings of EKAW 2002, Sigüenza, Spain, 2002.

⁴⁴Gruber, T. R. (1993a). *Toward principles for the design of ontologies used for knowledge sharing*. In Guarino, N. and Poli, R., editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Dordrecht, The Netherlands. Kluwer Academic Publishers. See also, Gangemi A, Pisanelli DM, Steve G: An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies. *Data and Knowledge Engineering*, 1999, 31, pp. 183-220, 1999.

⁴⁵Consider, however, that all the “lightweight ontologies” introduced in the following paragraphs (WordNet, Frame-Net and CYC) are the results of an intense, time-consuming human labour.

⁴⁶Masolo C., Borgo S., Gangemi A, Guarino N, Oltramari A. The WonderWeb Library of Foundational Ontologies, IST 2001-33052 Wonder Web.

<http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf>, 2003.

the foundational ontology by a Legal Core Ontology which describes the basic entities (legal roles, legal situations, duties, sanction, organizations, etc.) of the legal reality⁴⁷

It is possible that a lexicon (considered as a lightweight ontology) with a semantic hierarchy might serve as the basis for a useful ontology, and that an ontology may serve as a grounding for a lexicon. This is particularly the case in technical domains, in which vocabulary and ontology are more closely tied than in more general domains (Hirst 2003, p.14). In terminological lexicons, terms and concepts usually coincide, creating an intersection between linguistic meaning and formal ontological conceptual structure.

7.2 General Lexical Resources

7.2.1 WordNet

WordNet⁴⁸ is a lexical database which contains information about English nouns, verbs, adjectives and adverbs and it is organized around the notion of *synset*. A synset is a set of words with the same part-of-speech that can be interchanged in a certain context.

WordNet can be used for supporting semantic inferences, for finding alternative expressions or wordings, for expanding words to sets of semantically related or close words (in e.g. information retrieval), for automatic text analysis and artificial intelligence applications.

WordNet is free downloadable and it is maintained by the Cognitive Science Laboratory of Princeton University under the direction of professor G. A. Miller and Pr. C. Fellbaum. The database contains about 150,000 words organized in over 115,000 synsets for a total of 207,000 word-sense pairs. Every synset contains a group of synonymous words that go together to form a specific meaning; different senses of a word are in different synsets. The meaning of the synsets is further clarified with short defining glosses. Synsets are connected to other synsets via a number of semantic relations. These relations include: *hyp(er)onymy*, *troponymy*, *holo(mero)nymy*, *entailment* and *cause relation*.⁴⁹

Moreover, words can also be connected to other words through lexical relations, including *antonymy* and relations of *derivation*.

⁴⁷e.g. Rinke Hoekstra, Joost Breuker, Marcello Di Bello, and Alexander Boer. *The LKIF Core ontology of basic legal concepts*. In Pompeu Casanovas, Maria Angela Biasiotti, Enrico Francesconi, and Maria Teresa Sagri, editors, Proceedings of the Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007), June 2007. <http://www.estrellaproject.org/lkif-core>

⁴⁸Fellbaum C. (ed.), WordNet: An Electronic Lexical Database and Some of its Applications, MIT Press, 1998. www.wordnet.princeton.edu/

⁴⁹Moreover, a relation of coordination is given when the two terms X and Y share a hyperonym. As: holonym: Y is a holonym of X if X is a part of Y; meronym: Y is a meronym of X if Y is a part of X.

WordNet also provides the polysemy count of a word, i.e. the number of synsets that contain the word. A frequency score is also provided, indicating how frequent the specific meaning is in a sample of annotated sentences.

The Princeton WordNet is linked through the ILI (Interlingual Index) to the lexical/semantic networks developed, for European languages, within the EU EuroWordNet⁵⁰ project. The EuroWordNet multilingual lexical database consists of wordnets for eight European languages, which are structured along the same lines as the Princeton WordNet. Currently, more than twenty languages share the same methodology and development structure and are linked to each other through the English language.

7.2.2 FrameNet

The Berkeley FrameNet Project⁵¹ (Baker, Fillmore, & Lowe 1998; Fillmore & Baker 2001) is building an online lexical resource for English, based on the principles of Frame Semantics (Fillmore, 1968), on the notion of thematic role and supported by corpus evidence. A semantic frame is a script-like structure of inferences, which are linked to the meanings of linguistic units (lexical items). Each frame identifies a set of frame elements (FEs), which are frame-specific semantic roles. The description of each lexical item identifies the frames which underline a given meaning and the ways in which the FEs are realized in structures headed by the word. FrameNet includes a rich network of relations between frames, such as:

- Inheritance (IS-A relation): the child frame is a subtype of the parent frame, and each FE in the parent is bound to a corresponding FE in the child;
- Using: the child frame presupposes the parent frame as background;
- Sub frame: the child frame is a sub-event of a complex event represented by the parent.
- Perspective on: the child frame provides a particular perspective on an un-perspectives parent frame

The FrameNet database documents the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses, through manual annotation of example sentences and automatic summarization of the resulting annotations. FrameNet currently contains more than 10,000 lexical units, more than 6,100 of which are fully annotated, in more than 825 semantic frames, exemplified in more than 135,000 annotated sentences.

The FrameNet database is available in XML, and can be displayed and queried via the Web and other interfaces. The FrameNet I data has also been translated into the DAML+OIL extension to XML and the Resource Description Framework (RDF), which can represent our ontology and to make FrameNet information machine readable and understandable.

⁵⁰Ide N., Greenstein D., Vossen P. (eds.), Special Issue on EuroWordNet, in *Computers and the Humanities*, Volume 32, Nos. 2-3 1998, Kluwer Academic Publishers, Dordrecht, 1998.

⁵¹ <http://framenet.icsi.berkeley.edu/f/%ramenet>

7.2.3 CYC

Cyc⁵² is an artificial intelligence project that attempts to assemble a comprehensive ontology and database of everyday common sense knowledge, with the goal of enabling AI applications to perform human-like reasoning.

The Cyc project, started in 1984, is predicated on the idea that effective machine learning depends on having a core of knowledge that provides a context for novel learned information known as “common sense”⁵³. The purpose of the Cyc project is to provide computers with a store of formally represented “common sense”: real world knowledge that can provide a basis for additional knowledge to be gathered and interpreted automatically.

Over three million facts and rules have been formally represented in the Cyc knowledge base by ontologists skilled in CycL, Cyc’s formal representation language. In addition, natural language generation and parsing capabilities have been developed to provide support for learning from English corpora. As a result, the Cyc knowledge base now contains enough knowledge to support experimentation with the acquisition of additional knowledge via machine learning.

The Cyc system is made up of three distinct components: the knowledge base (KB), the inference engine, and the natural language system. The Cyc KB contains more than 3.2 million assertions (facts and rules) describing more than 280,000 concepts, structured into a hierarchical graph of micro theories, or reasoning contexts.

The natural language component of the system consists of a lexicon, and parsing and generation subsystems. The lexicon is a component of the knowledge base that maps words and phrases to Cyc concepts, whereas various parsers provide methods for translating English text into CycL, a declarative language used in the project.

The original knowledge base is proprietary, but a smaller version of the knowledge base, intended to establish a common vocabulary for automatic reasoning, was released as OpenCyc under an open source license. OpenCyc is the open source version of the Cyc <http://www.cyc.com/cyc/technology/whatis-cyc-technology>, the world’s largest and most complete general knowledge base and common-sense reasoning engine.

⁵²See the Cyc’s website: <http://www.cyc.com>

⁵³Matuszek C. ., Cabral J, Witbrock M., DeOliveira J. *An Introduction to the Syntax and Content of Cyc*. In Proceedings of the 2006 AAAI Spring Symposium on Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering, Stanford, CA, March 2006.

7.3 Lexical resources for legal domains

Semantically enriched vocabularies are valuable in aiding part-of-speech tagging, thus helping automatic semantic and syntactic analysis of texts, supporting conceptual and cross-lingual I.R., ensuring consistency in legislative drafting.

When building a legal vocabulary, we have to consider that two different types of semantic information are associated with elements from legal texts. On the one hand, there is ontological structuring in the form of a conceptual model of the legal domain. A legal ‘language’, consisting of a complex structure of concepts, forms an abstraction from legal textual material. On the other hand, there is a vocabulary of lexical items that lexicalise concepts (a lexicon), which are not necessarily restricted to the legal domain, and are associated with specific linguistic information (e.g. nouns versus verbs and syntactic preference). Variable dependency of ontological structure on language constraints indicates a continuum between formal structure and linguistic description. On the formal end of the continuum are maximally language independent ontology reflecting the structure of the domain in question. On the linguistic end there is a complete ontological dependency on domain-specific lexicons and grammars. This means that, in the latter case, the structure of the ontology is fully determined by the structure of the language. A knowledge base that models legal knowledge needs to take both types of information into account, and establish a modularly organized integration of the two.

In the following we describe the terminological sources created by the EU Institutions to support the multilingualism issues, and two projects developing semantic resources for legal domains.

7.3.1 The European Institutional terminological repositories: Eurovoc and Iate

The most prominent EU thesaurus is *Eurovoc*⁵⁴. Eurovoc is a multilingual thesaurus – a controlled vocabulary – covering the policy fields of the EU. It provides a means of indexing the documents in the documentation systems. The latest version - Eurovoc 4.2 - exists in 16 official languages of the European Union (Spanish, Czech, Danish, German, Greek, English, French, Latvian, Italian, Hungarian, Dutch, Polish, Portuguese, Slovene, Finnish and Swedish). Eurovoc has a hierarchical structure with inter-lingual relations. As the focus is on socio-economic issues, depth in law is quite low and the structure is not appropriate to EU law. The classification codes (or headings) of the *Register of the Community law in force* represent much higher quality for legal purposes. Inter-lingual relations exist and also a relatively fine-grained hierarchical structure is present. Depth is still not sufficient. The quite powerful *descriptors of the European Court of Justice* are more a list of legal sentences (“Rechtssätze”) than a proper thesaurus.

Eurovoc pertains to the category of traditional thesauri, structured on hierarchical and synonymy relations, with a rigid structure and a poor semantics. Eurovoc has a broader scope

⁵⁴eurovoc.europa.eu/

(European policy issues), and the components dedicated to the normative domain are very low in precision and granularity. As Eurovoc was drafted exclusively for manual indexing and retrieval purposes, the lack of semantic precision generates frequent inconsistencies among several hierarchical and synonymy relations so that it is mainly suitable for retrieving *related* terms. In the same way as all existing thesauri it is focused on documentation and lacks sufficient granularity for semantic access to EU law.

A second EU terminological resource is IATE (Inter-Active Terminology for Europe), the EU inter-institutional terminology database recently opened to public use⁵⁵. It generates from the integration of the Eurodicautom database, TIS, Euroterms and all the terminology produced by the Institutional Agencies and Translation Centers of the Eu Institutions (Commission, Parliament, Court of Justice, Court of Auditors). The aim of Iate is to support European translators in keeping track of translations of words and phrases in different languages, and in establishing potential candidates for the translations of certain terms. The innovative aspect of Iate is the web-based infrastructure enhancing the availability and standardisation of the information and the dynamic evolution of the database, which now contains approximately 1.4 million multilingual entries.

7.3.2 The TESEO - TEsauro SENato per l'Organizzazione dei documenti parlamentari

The TESEO thesaurus is the semantic tool used by the Italian Parliament to classify documents of databases collecting legislation and other relevant contents related to the legislative activity. It is a thesaurus able to describe the entire social reality. It is made up of 3661 descriptors organized according to hierarchical, equivalence and associative relationships. Citizens can access the Parliament databases navigating the thesaurus and choosing the most appropriate descriptor to be used in their search. TESEO has been adopted also at regional level by some local public authorities.

7.3.3 The LOIS Project

LOIS is a recently concluded Demonstration Project co-financed by the European Commission under the e-Content Program. The main objective of LOIS is the localization of WordNets describing the legal domain into six different European languages, namely Italian, English, German, Czech, Portuguese and Dutch. The synsets (or concepts) of national legal WordNets are linked across them, in such a way to guarantee cross lingual access to European legislation and other legal documents (such as court cases). Citizens and/or professional users then are enabled to enter queries to a legal documentation base in their language and retrieve also documents written in different languages. Localization has been carried out in such a way that cross-lingual relationships between different synsets are maintained⁵⁶.

⁵⁵<http://iate.europa.eu/>

⁵⁶Curtoni P., Di Tomaso V., Dini L., Mommers L, Peters W., Quaresma P., Schweighofer E., Tiscornia D., *Semantic access to multilingual legal information*, in: Workshop su "Free EU Information on the Web: The Future beyond the new EUR-Lex", Bruxelles, 2005

As its methodological starting point, LOIS adopts the structure of WordNet and EuroWordNet (EWN)⁵⁷. As in Princeton WordNet, in LoIs, a *synset* is a set of one or more uninflected word forms (lemmas) with the same part-of-speech that can be interchanged in a certain context. As already explained, for example the words {*case, cause, causa, law suit*} form a *noun synset* because they can be used to refer to the same concept. A synset is often further described by a gloss.

The LOIS database is compatible with the *EuroWordNet* architecture⁵⁸, and forms an extension of the EWN semantic coverage into the legal domain. Overall, LOIS consists of a number of modules that directly or indirectly link into EWN modules through each individual language component (see figure 2 for a simplified view on the database structure).

LOIS contributes to the creation of a European Legal WordNet, where lexical items may be mapped from language to language according to their semantic meaning, at least as far as we can consider the structure of WordNet as semantically oriented. The goal of LOIS, however, is not only to localize resources for the sake of exploitation of third parties. LOIS aims at creating a paradigm for accessing legal resources both from a monolingual and cross lingual point of view. The starting point is a legal WordNet which has been developed by ITTIG⁵⁹.

As already pointed out, one of the typical obstacles the citizen is faced with when accessing legal resources is represented by the so called “Legal Jargon”. Indeed several legal concepts may have different denominations in the common language and the legal language. By maintaining links between several Legal European Wordnets with the corresponding “standard language” wordnets (whenever available), the localized resources provide a formal way to map standard language onto legal jargon and vice versa.

The Information Retrieval engine which is at the base of LOIS, is able to exploit such links, in a way that the query of a user unable to use technical terms, retrieves also documents which contain technical terminology.

Even professional people in the legal field might have problems in retrieving certain kinds of legal documents by using appropriate keywords. This emerges with a particular strength when they have to look for case law (judicial decisions), where the discursive style of the argumentation is affected by the semantic ambiguity of natural language. The same legal concepts can indeed be designed with different terms which are just synonyms from a legal

⁵⁷See paragraph 7.1

⁵⁸Within the LOIS database, synsets are related to each other by means of semantic relations, of which the most important are hypernymy/hyponymy (between specific and more general concepts), meronymy (between parts and wholes), and antonymy (between semantically opposite concepts); even if less used, LOIS also includes all EWN relations

⁵⁹Sagri M. T., Tiscornia D, *Semantic Tools for Accessing Legal Information*, A. Palma dos Reis, P. Isahas (a cura di), Atti del Convegno IADIS: E-Society 2003 (Lisbona, 3-6 Giugno 2003), IADIS press, Lisbona, 2003, pp. 685- 688

point of view. Automatic query expansion on the base of synsets will lead to a better user satisfaction in case of unresponsive queries (i.e. queries by which fewer documents than expected are retrieved).

Currently, the LOIS database covers six legislative systems coinciding with six languages. In line with the discussion of legal language above, each LOIS national legal wordnet is composed of two types of database modules:

- an indigenous *lexical database*, which conceptualizes general language entities pertaining to legal theory and legal dogmatism, a set of patterns (models) in line with which law is formed and operates, and which is structured according to the EWN methodology;
- a *legislative database*, populated by legal concepts defined in European and national legislation and structured according to purely legal (supra)national models⁶⁰.

In order to connect linguistic expressions of concepts to the underlying conceptual domain entities, an intermediate structure (a *core ontology*) is needed, made up of units of understanding, to distinguish language-independent concepts and relations from concepts and relations which are not. A *core legal ontology* is a complete and extensible ontology that expresses the basic concepts of Law, and that can provide the basis for specialization into domain-specific concepts and vocabularies. A *core legal ontology* such as *CLO*⁶¹ and *LRI-Core*⁶² intends to bridge the gap between domain-specific concepts and the abstract categories of formal upper level or foundational ontologies such as DOLCE⁶³, transforming lexical relations in formal properties consistent with the top-down formal semantics imposed by the upper ontology. Given a superset of entities and relations in six WordNets, each pertaining to a particular legal system, this logical backbone will help to distinguish language-independent and language-dependent concepts and relations.

7.3.4 The Syllabus Project

One tool which might help to increase European terminological consistency taking into account the problem of knowledge representation is the Legal Taxonomy Syllabus (Syllabus) which has already been set by the Department of Computer Science of the University of Turin specialized in ontologies for the web domain, legal ontologies and AI studies in law. Such a

⁶⁰Peters W., Sagri M. T., Tiscornia D., Castagnoli S., The LOIS Project, *The fifth international conference on Language Resources and Evaluation*, LREC, Genova, 2006

⁶¹Gangemi A., Prisco A., Sagri MT, Steve G, Tiscornia D, *Some ontological tools to support legal regulatory compliance, with a case study*, in Jarrar M, et al. (eds.), *Proceedings of the WORM03 Workshop at OTM Conference*, Springer Verlag, Berlin, 2003.

⁶²Breuker, J.A., Valente, A. & Winkels, R.. Use and reuse of legal ontologies in knowledge engineering and information management. In Benjamins e.a., editors, *Law and the Semantic Web*, pp. 36-64. Springer Verlag, Berlin. Volume 3396, 2005.

⁶³Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L. (2002), Sweetening Ontologies with DOLCE. In: *Proceedings of EKAW 2002*.

tool has been developed to support the Uniform Terminology project⁶⁴ in the overall collaboration with the Law Department of Turin⁶⁵.

Legal Taxonomy Syllabus is designed as an open-access database linking European terms with national transposition law and also linking terms horizontally (i.e. between national legal orders). It provides full text reference of relevant EU and Member States' legislation.

The database includes related case law and short commentary notes by national scholars where this is necessary to describe differing legal doctrine. As a starting point, the Legal Taxonomy Syllabus covers consumer law with national law references limited to France, Germany, Italy, Spain and the UK.

The Legal Taxonomy Syllabus could be useful for lawyers, translators, legislators and scholars: the tool may help legislators to enhance terminological coherency already at drafting stage of legal acts.

The cross-reference features enable lawyers to search for relevant case law in other Member States by a one-click method. Unlike a dictionary, the Legal Taxonomy Syllabus does not only provide translators with a proper translation but provides the respective legislative context of each term. The database can also help scholars to perform cross-sector analyses on the use of legal terms and concepts, integrating with the ontologies the classification systems employed by existing database, which are oriented towards the multilingualism terminology but only in one dimension, that of the European Union; or which are structured on the subdivisions and sections of the European Treaty that have no relevant connection with the classifications implied in the legal discourse within the Member States.

The treatment of legal terminology adopts a mixed descriptive-prescriptive terminological application to the corpora constituted by the European legal documents. In this way, prior to improving the consistency of the EU terminology, the research is focused on the understanding of such a terminology within the Member States, highlighting the matter of polysemy and amphibology of the terms used in a same language at the European level and at the national level.

The tool is based on a clear distinction between the notions of “legal term” and “legal concept”. The basic idea is that the basic conceptual backbone consists in a taxonomy of concepts (lightweight ontology) to which the terms can refer to express their meaning. One of the main points is that, in the legal domain, there is not only a single taxonomy covering all languages. Consequently, the Legal Taxonomy Syllabus includes different ontologies, one for each involved language plus one for the language of EU documents. Each language-specific ontology is related via a set of “association” links to the EU concepts.

⁶⁴P. Rossi and C. Vogel. *Terms and concepts; towards a syllabus for european private law*. European Review of Private Law (ERPL), 12(2):293–300, 2004.

⁶⁵<http://www.copec1.org>

7.3.5 The DALOS Project

The DALOS project, financed under e-Participation program, aims at providing legal drafters and decision-makers with linguistic and knowledge management tools to be used in the legislative processes, in particular within the phase of legislative drafting. This will contribute to the harmonization and coherence of legislative texts by means of:

- providing the drafters with easy access to the pertinent ontological resources so that they can obtain immediate clarifications on legal terms used, navigate through the network of legal terms, consider the ways in which the same concept is expressed in different languages;
- enriching linguistic-ontological resources both via interaction with legal experts directly involved and via the integration of ontology learning tools, able to cluster relevant terms from legislative texts and organizing them in taxonomical chains integrating ontological tools within the legislative process;
- facilitating accessibility to legislative texts by European citizens, thus promoting their participation in the legislative process and their involvement in the law comprehension process;
- improving the quality and the readability of legislative texts, thus contributing also to the “certainty of law”.

Moreover, once the text is delivered, ontological knowledge will facilitate retrieval for European citizens and also interpretation and implementation of legal documents. These results, besides a well structured multilingual domain-specific ontology require a set of computer tools, clustered around such ontology. Such tools will allow to monitor the connection between the language and terminology used in the text and the harmonized concept repository, thus allowing the law-makers to get immediate feedback on the quality and accessibility of the language used. On the other hand, tools for terminology extraction and ontology learning can help the expansion of existing ontology and terminological repositories in a controlled and harmonized way. In DALOS project such tools are expected to manage multilingual issues: they are addressed to harmonize legal concepts and related terminologies used in European legislation as well as in the legislation of different European countries. To obtain this result the project is exploiting linguistic ontological resources developed within previous European project experiences. In particular to guarantee the feasibility of the project, the ontological-terminological resources developed within LOIS project⁶⁶, are made available for their integration within the drafting tools, used by legislative offices, through a defined standard interface, thus without interfering in the chain of legislative production, but nevertheless guaranteeing the improvement of the law-making activity and the final products of the legislative processes. Using such resources legislative drafters will be able to query linguistic and ontological resources, searching for appropriate and standardized terms or locutions, corresponding to specific legal concepts. The above resources will be integrated with T2K (“Text-to-Knowledge”), an ontology learning tool jointly developed at CNR-ILC and Pisa University (Department of Linguistics) which combines linguistic and statistical techniques to carry out the ontology learning task. Starting from a document collection, T2K acquires domain terminology (both simple and multi-word terms) and organizes the set of acquired terms into

⁶⁶ EDC 22161- see <http://www.loisproject.org>

a) taxonomical chains (reconstructed from their internal linguistic structure), and
b) clusters of semantically related terms (inferred through distributionally based similarity measures). Moreover the GATE tool developed by the University of Sheffield Computer Science Department will be integrated as well. It supports advanced language analysis, data visualization, and information sharing in many languages.

GATE has facilities for viewing, editing, and annotating corpora in a wide number of languages (based on Unicode) and has been used successfully for the creation, semi-automatic annotation and analysis of many electronic resources. DALOS prototype is focused on the specific area of protection of consumers' economic and legal interests.

The final result of DALOS project will be a linguistic tool assisting the process of drafting legislation by providing ontology-based and statistically selected suggestions for terminological re-use and knowledge management.

7.3.6 The Euroterm Project (EDC-2214)

Due to the difficulties multi-languages poses for the EU Institutions and Member States in all sectors of Public Sector Information, EuroTerm⁶⁷ ("Extending the EuroWordNet with Public Sector Terminology"⁶⁸) is essentially dedicated to extending EuroWordNet with environmental terminology for the following languages: Greek, Dutch and Spanish, by adding specific terms of the public sector. Therefore, this project allows the creation of a multi-language lexicon database to favour access to information and to contribute to the use and development of European digital contents.

Each of the monolingual WordNets comprises of ~1,000 synsets, stored in a common database, which will be linked to the central EWN database under the domain label "Environment".

7.4 Semantic Approach Trends

Existing classification schemes and thesauri are lacking in well-defined semantics and structural consistency. Empowering end users in searching collections of ever increasing magnitudes with performance far exceeding plain free-text searching (as used in many Web search engines), and developing systems that not only find but also process information for action, requires far more powerful and complex knowledge organization systems such as ontologies (lightweight and foundational) which allows to achieve the biggest challenge in Information retrieval that is concept identification in specific domain.

The use of ontologies allows to share common understanding of the structure of information among people or software agents which is one of the more common goals in developing ontologies (Musen 1992; Gruber 1993). For example, suppose several different Websites

⁶⁷Christodoulakis D., *Extending EWN using both the expand and merge model*, in Jarrar M, et al. (eds.), Proceedings of the WORM03 Workshop at OTM Conference, Springer Verlag, Berlin.

⁶⁸<http://www.ceid.upatras.gr/Euroterm/>

contain legal information or provide legal e-commerce services. If these Websites share and publish the same underlying ontology of the terms they all use, then computer agents can extract and aggregate information from these different sites. The agents can use this aggregated information to answer user queries or as input data to other applications. Then ontology enables reuse of domain knowledge which was one of the driving forces behind recent surge in ontology research. For example, models for many different domains need to represent the notion of time. This representation includes the notions of time intervals, points in time, relative measures of time, and so on. If one group of researchers develops such ontologies in detail, others can simply reuse it for their domains. Additionally, if we need to build a large ontology, we can integrate several existing ontologies describing portions of the large domain.

Furthermore, using ontology is possible to make domain assumptions explicit underlying an implementation and changing these assumptions easily if the knowledge about the domain changes. Hard-coding assumptions about the world in programming-language code makes these assumptions not only hard to find and understand but also hard to change, in particular for someone without programming expertise. In addition, explicit specifications of domain knowledge are useful for new users who must learn what terms in the domain mean.

Separating the domain knowledge from the operational knowledge is another common use of ontologies.

Finally, the use of ontologies in a context such as the legal one seems to be relevant due to the fact that Law appears not ontologically founded.

References

- Africa i-Parliament, and UNITED NATIONS Department of Economic and Social Affairs. *AKOMA NTOSO Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontologies*, 2007.
- AIPA. 2001, (November). Definizione delle regole per l'assegnazione dei nomi uniformi ai documenti giuridici. Circolare n. AIPA/CR/35. In Italian
- AIPA. 2002, (April). Formato per la Rappresentazione Elettronica dei Provvedimenti Normativi tramite il Linguaggio di Marcatura XML. Circolare n. AIPA/CR/40 (in Italian), http://www.cnipa.gov.it/site/_contentfiles/00127500/127544_CR_40_2002.pdf.
- Alchourryn, C. E., P. Gärdenfors, and D. Makinson. 1985. On the Logic of Theory Change: Partial Meet Functions for Contractions and Revisions. *Journal of Symbolic Logic* 50: 510–30.
- Arnold-Moore, T. 1997. Automatic generation of amendment legislation. In *Proceedings of the International Conference of Artificial Intelligence and Law*.
- Arnold-Moore, T., J. Clemes, and M. Tadd. 2002. Connected To The Law: Tasmanian Legislation Using EnAct. Technical report, TeraText.
- Benkler, Y. 2006. *The Wealth of Networks: How Social Production transforms Markets and Freems*. New Haven, Conn.: Yale University Press.
- Biagioli, C. 1997. Towards a legal rules functional micro-ontology. In *Proceedings of workshop LEGONT '97*.
- Biagioli, C., A. Cappelli, E. Francesconi, and F. Turchi. 2006. Law Making Environment: perspectives. In *Proceedings of the V Legislative XML Workshop*. To appear
- Bing, J. 1984. *Handbook of Legal Information Retrieval*. Amsterdam: North Holland.
- Bing, J. 2003. The Policies of Legal Information Services: A Perspective of Three Decades. In *Yulex 2003*. Ed. L. A. Bygrave, 37–55. Oslo: Norwegian Research Centre for Computers and Law.
- Bing, J., and M. G. Schoenberg. 1994. Improving Regulatory Management: The Use of Information Systems. NORIS (96) II (Fourth version)
- Boer, A., R. Hoekstra, and R. Winkels. 2002. MetaLex: Legislation in XML. In *Proceedings of JURIX 2002: Legal Knowledge and Information System*, 1–10.
- Boer, A., R. Winkels, R. Hoekstra, and T. van Engers. 2003. Knowledge Management for Legislative Drafting in an International Setting. In *Proceedings of JURIX 2003: Legal Knowledge and Information System*, 91–100.
- Boer, A., R. Winkels, and F. Vitali. 2007. XML Standards for Law: MetaLex and LKIF. In *Proceedings of JURIX 2007*. Amsterdam: IOS.
- Borruso, R. 1978. *Civilta del computer*. Milano: IPSOA.
- Castells, M. 2000. *The Rise of the Network Society*. Oxford: Oxford University Press.
- Castells, M. 2001. *The Internet Galaxy*. Oxford: Oxford University Press.

- Commission of the European Communities. 2002. Better Lawmaking 2002. A shared Responsibility, Commission Report to the European Council, COM(1998) 715 final.
- Commission of the European Communities. 2006. A strategic review of Better Regulation in the European Union Commission communication -COM(2006)689 (14 November 2006).
- Dayal, S., and P. Johnson. 1999. A Web-based Revolution in Australian Public Administration. In *Proceedings Law via the Internet '99: 2nd AustLII Conference on Computerisation of Law via the Internet*. Sydney: University of Technology. (Electronic publication. Also Published in the online journal JILT, 2000, Issue 1.)
- EU. 2003. Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information.
- Free Access to Law Movement. 2004. Declaration on Free Access to Law.
- Goodin, R. 2003. *Reflective Democracy*. Oxford: Oxford University Press.
- Gordon, T. F. 1995. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Dordrecht: Kluwer.
- Gordon, T. F. 2007. Constructing Arguments with a Computational Model of an Argumentation Scheme for Legal Rules. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law (ICAIL-2007)*, 117–21. ACM.
- Gore, A. 2007. *The Assault on Reason*. New York, N. Y.: Penguin.
- Governatori, G., A. Rotolo, R. Riveret, M. Palmirani, and G. Sarto. 2007. Variants of temporal defeasible logics for modelling norm modifications. In *Proceedings of Eleventh International Conference on Artificial Intelligence and Law*, 155–9. New York, N. Y.: ACM.
- Greenleaf, G. 2004. Jon Bing and the History of Computerised Legal Research: Some Missing Links. In *Et tilbakeblikk pa Fremtiden*. Ed. O. Torvund and L. A. Bygrave, 61–75. Institutt for rettsinformatikk.
- Greenleaf, G., Mowbray, G. King, and P. van Dijk. 2002. Public access to law via Internet: the Australasian Legal Information Institute.
- Hage, J. C. 1997. *Reasoning with Rules: An Essay on Legal Reasoning and Its Underlying Logic*. Dordrecht: Kluwer.
- ICAIL. 2007. *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*. New York, N. Y.: ACM.
- Khaled, O. A., H. Chabbi, and M. Ramalho. 2004, July). *ĪTUDE URN*. University of Applied Sciences of Western Switzerland, EIA-FR
- Leith, P., and A. Hoey. 1998. *The Computerised Lawyer*. London: Springer.
- Lupo, C., and C. Batini. 2003. A Federative Approach to Laws Access by Citizens: The “NormeinRete” System. In *Proceedings of Second International Conference Electronic Government (EGOV)*. Ed. R. Traunmuller, 413–416. Springer-Verlag.
- Maeda, M. 2003. *Workbook Hosei Shitsumu*. (revised ed.), Gyosei, Tokyo
- McClure, J. 2006. Legal-RDF Vocabularies, Requirements & Design Rationale. In *Proceedings of the V Legislative XML Workshop*. To appear
- Megale, F., and F. Vitali. 2001. I DTD dei documenti di Norme in Rete. *Informatica e Diritto* 1: 167–231.

- Moats, R., and K. R. Sollins. 1997. URN Syntax. Technical Report RFC 2141, Internet Engineering Task Force (IETF).
- OCDE. 1994. Improving the Quality of Laws and Regulations: Economic, Legal and Managerial Techniques. OCDE/GD(94)59. Technical report.
- OCDE. 1997. Regulatory Impact Analysis: Best Practice in OECD countries, OECD, Paris.
- Ogawa, Y., S. Inagaki, and K. Toyama. 2007. Automatic Consolidation of Japanese Statutes based on Formalization of Amendment Sentences. In *Proceedings of the First International Workshop on Juris-informatics*.
- Palmirani, M., and R. Brighi. 2003. An XML Editor for Legal Information Management. In *Proceedings of DEXA 2003, the 2nd International Conference on Electronic Government - EGOV 2003*, 421–9. Berlin: Springer.
- Palmirani, M., and R. Brighi. 2006. Time Model for Managing the Dynamic of Normative System. In *Electronic Government, 5th International Conference, EGOV 2006, Krakow, Poland, September 4-8, 2006*. Ed. M. Wimmer, H. J. Scholl, E. Grunlund, and K. V. Andersen, 207–18. Springer.
- Palmirani, M., R. Brighi, and M. Massini. 2005. Automated Extraction of Normative References in Legal Texts. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, 105–6. New York, N. Y.: ACM.
- Prakken, H. 1997. *Logical Tools for Modelling Legal Argument: A Study of Defeasible Reasoning in Law*. Dordrech: Kluwer.
- Prakken, H., and G. Sartor. 1997. Argument-based Extended Logic Programming with Defeasible Priorities. *Journal of Applied Non-classical Logics* 7: 25–75.
- Sartor, G. 2005. *Legal Reasoning: A Cognitive Approach to the Law*. Volume 5 of *Treatise on Legal Philosophy and General Jurisprudence*. Berlin: Springer.
- Sartor, G. 2006. Fundamental Legal Concepts: A Formal and Teleological Characterisation. *Artificial Intelligence and Law* 21: 101–42.
- Sartor, G. 2007. The Nature of Legal Concepts: Inferential Nodes or Ontological Categories. Working paper, European University Institute, Florence.
- Simitis, S. 1977. *Crisi dell'informazione giuridica ed elaborazione elettronica dei dati*. Milano: Giuffrè.
- Spinosa, P. 2001. Identification of Legal Documents through URNs (Uniform Resource Names). In *Proceedings of the EuroWeb 2001, The Web in Public Administration*.
- Tucker, H. 2004a, July). *Lex Dania - White Paper. A System of XML Schemas for Danish Legislative Documentation*.
- Tucker, H. 2004b, July). *LexDania - Documentation. Guidelines for Writing Omni- and DocType and Application Schemas*.
- Tucker, H. 2004c, March). *LexDania - Documentation. Introduction to Concepts and Status of XML Linking*.
- Tucker, H. 2004d, June). *LexDania - Documentation. Metadata*.
- Tucker, H. 2004e, May). *LexDania - Documentation. XML and XSL Best Practices for Writing Schemas*.

- Tucker, H. 2004f, April). *LexDania Documentation. Universal addresses*.
- van Engers, T. M. 2006. *Proceedings of the Nineteenth Annual Conference on Legal Knowledge and Information Systems (JURIX)*. Amsterdam: IOS.
- Vitali, F., and F. Zeni. 2007. Towards a Country-Independent Data Format: The Akoma Ntoso Experience. In *Proceedings of the V Legislative XML Workshop*, 67–86. European Press Academic Publishing.
- W3C. *Web Services Glossary: W3C Working Group Note 11 February 2004*, 2004.
- Waldron, J. 1999. *The Dignity of Legislation*. Cambridge: Cambridge University Press.