Published in F. Galindo & G. Quirchmayer (Eds.), *Proceedings of the IFIP 8.5 Working Conference on Advances in Electronic Government* (pp. 121-132). Zaragoza: Seminario de Informática y Derecho, Universidad de Zaragoza.

# Challenges for European legal information retrieval

VIRPI LYYTIKÄINEN, PASI TIITINEN, AIRI SALMINEN

*University of Jyväskylä*
*Department of Computer Science and Information Systems*
*P.O. Box 35, Fin-40351 Jyväskylä, Finland*

**Abstract**: Internationalisation and the expansion of the European Union have created a situation where the need for legal information from foreign countries and from different legislative levels is greater than ever before. This information is increasingly becoming available in digital form due to the increase in legal information repositories on the Internet. In spite of this, users of European legal information encounter multitude of problems in information retrieval. The information is scattered in numerous databases in which documents are organised and classified in different ways, their contents are written in different languages, and the functionality and user interfaces of the databases differ. Moreover the differences in legal systems greatly hinder the use of European legal information. The problems related to the retrieval of legal information in Europe have been studied in a project called EULEGIS. The main purpose of the project is to offer a consistent user interface to retrieve legal information created in different legal systems and at different levels in Europe – the European Union, a member state, a region, or a municipality. In the EULEGIS interface, information about the context of legal documents is given to users by graphical data models. In the paper we discuss the problems related to the complexity of the legal domain and solutions developed in the EULEGIS project.

**Keywords**: legal information retrieval, legal systems, visualisation, graphical overviews, legal metadata

# 1. INTRODUCTION

The new global information infrastructure has offered new means for extending the availability of recorded legal information. Three major groups of documents offering legal information can be identified: normative documents including the acts and decrees, preparatory works created during legislative processes, and court judgements showing how the rules have been interpreted. In many European countries, for example, in Norway, Denmark, Sweden, United Kingdom, and Finland, normative documents are available free-of-charge on the Internet. In addition, many fee-based services offer legal information on the Internet.

While the means for accessing information about the rules and regulations concerning people in Europe have extended, the set of the rules has also extended. In Finland, for example, joining the European Union in 1995 meant that all Finnish people and organisations became regulated not only by the national rules but also by the European Union rules. Internationalisation in general has also lead to an increasing need to know more about legal norms and regulations in other countries. In many situations, people in Europe need information created in different legal systems and at different levels – European Union, a country, a region, or a municipality. The need is obvious both among the people creating legal rules and among people trying to work and live according to the rules. For example, legislators in an EU member state have to be aware both of the European Union regulations and laws and legislative activities in other countries. A family planning to move to another European country may need lot of information related to regulations concerning schools, day care, insurances, pensions, and taxes. People planning business in European companies have to be aware of the EU rules, the rules in the countries where they operate as well as the forthcoming rules probably affecting their business in the future. The information needed may be scattered in several databases in which documents are organised and classified in different ways, their contents are written in different languages, and the retrieval techniques and user interfaces also differ. Finding the needed information is often hard even to information specialists and legal experts and may cause lot of expenses.

The problems related to the legal information retrieval in Europe have been studied in a project called EULEGIS (European User Views to Legislative Information in Structured Form) belonging to the Telematics Application Programme of the European

Commission. In the project nine partners from six European countries act together as a consortium aiming to develop new means for the retrieval of legal information. Many of the problems and challenges concerning the retrieval of European legal information are related to issues of complexity and heterogeneity in three areas: legal systems, digital legal information sources, and the users of the legal information. In the paper we will discuss these problem areas in Sections 2, 3, and 4, respectively. To help the European people to cope with the complexity of the legal domain in Europe, to better understand the differences in legal systems, and to better locate information from correct sources, a visual user interface offering important metadata in a uniform way has been developed in the EULEGIS project. The major parts of the interface are introduced in Section 5. Problems and issues faced in the development of the interface are discussed in the last section.

## 2. LEGAL SYSTEMS IN EUROPE

In democratic societies, the control and management of the society is based on documented public law, born under democratic legislative processes, and archived such that reliable validation for the law can be assured (Kiuru, Salminen & Chen, 1999). In spite that the laws and other legal rules created in Europe are all documented, knowing the legal rules concerning a person or an organisation is far from simple. The set of legal rules governing an entity is called a *legal system* or *legal order*. The entity in question can be, for example, a state, a group of states, an international organisation, a region, or a city. Citizens of any of the European countries are part of several coexisting and interacting legal systems. In Europe, the relationships of legal systems have recently been changing by the transfer of some of the legislative power previously exercised independently by the member states to the Union.

All legal systems are different in detail, although they may have some resemblance to each other. When the valid legislation and the role of the courts in the legislation is examined in different countries, two general types of legal systems can be identified: the common law system and the civil law system (see e.g. Gallagher, Laver & Mair, 1995). Within Europe, the jurisdictions of the United Kingdom and Ireland have a legal system based on common law. Most European countries belong to the civil law tradition, where the role of the case law has much less significance.

Another major difference in the legal systems of different countries is the differing role of regions. In some countries (e.g., Germany, Austria and Belgium), the legislative power has been decentralised to several power centres which have their own legal systems. On the other hand, there are countries only with minor legislative power at regional level. For example, in Finland there is only one region, which has an autonomous status and thus its own parliament. Certain amount of regulative power has often also been left, e.g., to provinces or municipalities.

## 3. LEGAL INFORMATION SOURCES

The various types of legal information sources and their significance depend greatly on the individual legal system. The three major groups of documents – normative, preparatory works, and court judgements – can be identified in all legal systems. The types of documents, their structures, and the languages used in them are, however, different. Even if the same language were used, the terminology could vary considerably.

As discussed in the introduction, the number of legal information providers on the Internet has been growing rapidly. Legal Web sites in Europe usually contain documents from one legal system only. Some databases specialise in one or more subject areas or contain only selected types of legal rules. Also, temporal coverage may vary among databases; some contain only recently published rules, while others may include legal norms dating back to the 18$^{th}$ century.

National legal documents are in many countries disseminated free-of-charge on the Internet (e.g. in Finland http://finlex.edita.fi, Sweden http://www.rixdagen.se/debatt/ or Norway http://www.lovdata.no). The dissemination of normative and preparative documents on the Internet free-of-charge has become policy in many countries. In addition, many fee-based services offer national legal information. Legal information concerning the European Union is also available at many Web sites, for example, CELEX (http://www.europa.eu.int/celex).

A Web site can offer several ways of accessing the legal information it contains. Legal rules may be accessed by date of publication, type of document, the organisation which originated the document, or subject. However, subject classification methods differ between Web sites. Also the functionality and presentation offered to the users of

these databases vary, and even the most frequently used functions may have been implemented in different ways. Moreover different fonts, colours, and layout make the casual use of multiple databases problematic.

In computerised information retrieval, the heterogeneity in the types, structures, languages, and terminologies of documents is extremely problematic. Even within one legal system and within documents of a specific type, inconsistencies have brought up problems not known at the time documented legal information was in printed form only. In reading printed documents, small varieties in the ways people have written some phrases or words, or varieties in the ways they have structured documents rarely cause any problems. In computerised information retrieval, even small inconsistencies may be a source of serious problems. The inconsistencies and their consequences in the retrieval situation are rarely well known to the user, and thus the way the person formulates the query easily leads to unreliable retrieval results. In the Finnish legal documents, for example, references to laws have two parts: the name of the law and the identifying number. People typing the text of legal documents have used to express the references in slightly different ways. The difference does not matter to a human reader. In some computerised retrieval systems however the differences affect the retrieval results. Kiuru (1994) identified in a database four different forms in the names of laws and five different forms in the numbers. Thus there were 4 * 5 = 20 alternatives to take into account when searching for a reference to a specific law in the database.

There are two major ways to solve the inconsistency problems discussed above. One is the deployment of more intelligent indexing schemes in legal databases. The retrieval system would then recognise, for example, a reference to a law written in slightly different ways. Research on developing more intelligent retrieval capabilities especially for the Web is an active research area. Another way to solve problems related to inconsistencies is standardisation, i.e., agreement upon more restrictive and more specific rules concerning the ways information is written in documents. SGML (Standard Generalized Markup Language, Goldfarb, 1990) is an international standard used for defining rules for document structures and for representing digital documents according to the rules. SGML-based standardisation of legal documents has taken place, for example, in Norway, Great Britain, and Finland.

## 4. USERS OF LEGAL INFORMATION IN EUROPE

All people need legal information but there are several particular user groups to whom foreign and EU legal information seems to be especially important. A problem in designing retrieval capabilities is that the experience and expertise of the users varies widely, both concerning the legal domain and the use of computers for information retrieval. Furthermore, the tasks in which the information is needed vary greatly, as well as the frequency of the needs. In some situations, rather general information is most useful while in other situations details may be important. An important user group consists of information intermediaries who help other people in finding and interpreting information.

User needs interviews related to legal information were carried out as part of the EULEGIS project and its Finnish predecessor RASKE during 1996-1999. RASKE was a project in which SGML-based document standards were developed for use in the production of documents in the Finnish Parliament and ministries (Salminen, Lehtovaara & Kauppinen, 1996, Salminen, Kauppinen & Lehtovaara, 1997, Salminen, 2000, Salminen, Lyytikäinen & Tiitinen, 1999). In-depth interviews were arranged in Finland and revealed that several groups of people regarded the accessibility of foreign legal information and EU legal information as important. Examples of these groups are people involved in national legislative processes, working in international and large companies, law firms, small and medium sized companies doing business with foreign partners, law institutions, and the mass media, as well as researchers and public administrators in general. The groups include both those highly expert in legal information retrieval as well as laymen. Legal information is also becoming more and more important for ordinary citizens who may, e.g., be changing their country of residence or buying property or goods from another EU state. In general, the need for foreign and EU legal information has been growing rapidly.

In spite of the differences in the interviewed user groups, in several of them similar needs were identified. The differences in various legal systems of Europe were mentioned as a major problem by the interviewed people. These differences often hindered both the search and the use of foreign legal information. Even where the access to the relevant documents had been gained, the utilisation of the documents was difficult without sufficient knowledge of the particular legal system. This was a

common problem especially for information specialists working in libraries and other public services whose clients regularly ask about legal situations in various European countries.

Other major problems were caused by the inconsistency and heterogeneity of different retrieval systems. Since the coverage of a single database was often insufficient to satisfy the information needs, several databases with varying frequency were needed. In cases where a database was not used regularly, it was often difficult to remember how to use its capabilities. In these situations the frequent changes in user interfaces caused extra harm.

## 5. VISUALISING LEGAL SYSTEMS FOR INFORMATION RETRIEVAL

The main purpose of the EULEGIS project has been to design a system with a consistent interface and consistent retrieval techniques, which can be used to retrieve legal information concerning different countries and different levels of legislation. The project aims to build a single access point for retrieving information from many of the differing databases that are available as Web resources. The need for such system is obvious but there are also lots of challenges and problems related to the design and implementation of such system. In planning the system, the reality of different and heterogeneous repositories has been accepted. An approach for improving the information retrieval from the distributed heterogeneous repositories would be the increase of intelligence in the retrieval techniques. The knowledge about the differences in the legal systems, legal documents, and needs of users would be stored and usable by the retrieval algorithms. This approach requires lot of research and implementation resources. In EULEGIS we have chosen another approach.

To help the European people to manage the complexity, to better understand the differences on the legal domain, and for being able to locate information from correct resources in spite of the differences and complexity, a visual user interface offering important metadata in a uniform way has been planned. The graphical interface also aids users in choosing the appropriate EULEGIS search forms for querying the documents.

In the EULEGIS interface, legal information retrieval is associated with legal systems. In spite that legal systems vary, they all have the following characteristics:

- They have *actors* (such as the Parliament or Ministry of Agriculture)
- They include different rules, which are found in *legal information sources* as different document types (e.g. Statutes or Government Bills)
- They implement different *processes* (e.g. preparation of a Government Bill).

Based on these common characteristics of legal systems, three different views to a legal system have been designed. The actor view describes the most significant actors of a certain legislative system, the legal source view shows the grouping of different kinds of legal documents, and the process view describes activities related to the legislative process.

The users can in many situations limit their query by the help of the views to concern only documents that are of a certain type, originate from a certain actor or from a certain phase of the legislative process. Views can describe, for example, the legal system of

- the European Union,
- a country (e.g. Finland, Sweden, United Kingdom or France),
- a region (e.g. the German state of Bavaria), or
- a municipality.

**Actor view**

The actor view describes the most significant actors that create the legal sources related to a legal system. Actors can be put into hierarchic actor groups shown graphically by nested rectangles. The actors belonging to an actor group may be part of a larger organisation (e.g. Parliament) or may otherwise have similar roles in the creation of the legal sources in question (e.g. different courts). The broken arrows are labelled by identifiers with phrases briefly describing the tasks of the actors or actor groups in the bottom part of the graph. An example of the actor view is shown in Figure 1.

By clicking the name of an actor the user can obtain more information about the role of the actor. From this additional information a link leads to a search form by which the user can search for the documents that originate from the selected actor (Figure 2).

**Legal information source view**

The legal information source view shows the most significant legal information sources in a legal system and also their relationships with each other. The graphical presentation of the view is composed of document types (i.e. legal information sources) and document type groups.

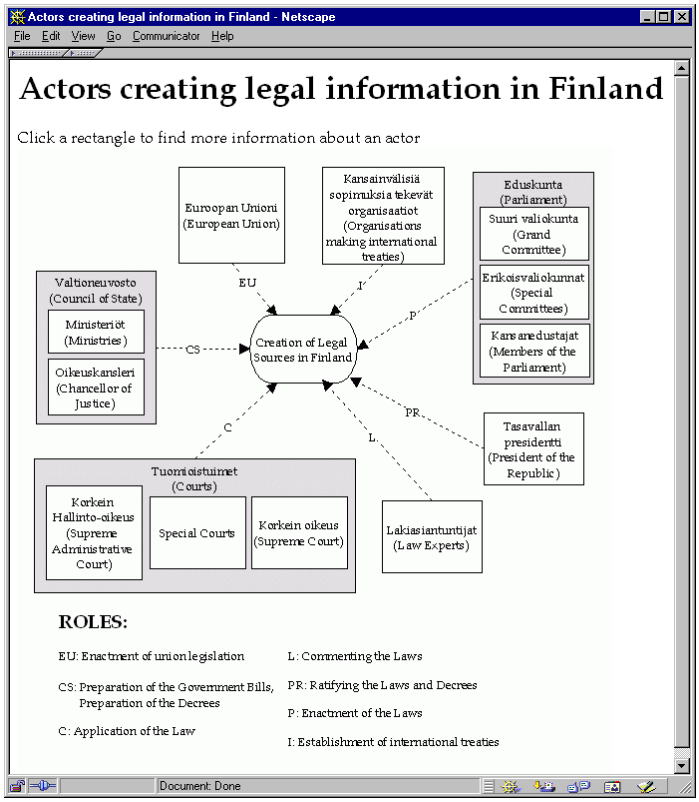


**Figure 1. The actor view to the Finnish legal system**

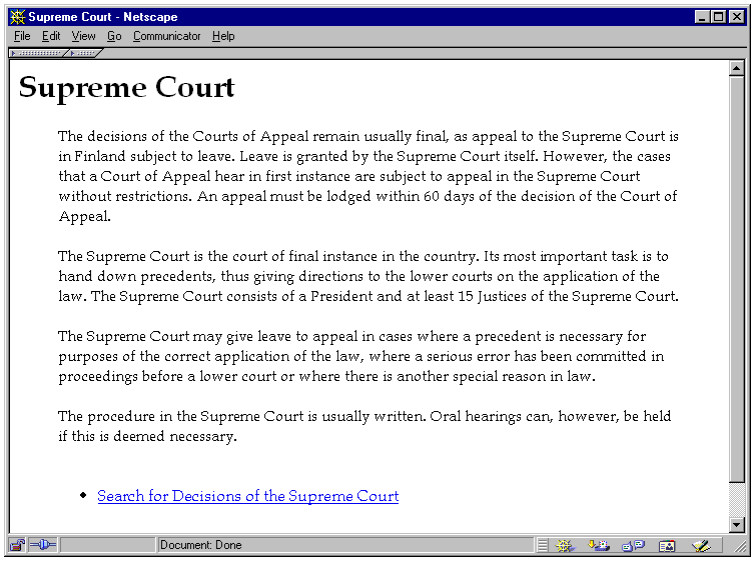


**Figure 2. Information about Supreme Court**

Figure 3 describes the legal information sources of the Finnish legal system as an example. National legal documents are categorised in five categories that exist in almost every country: preparatory works, normative documents, court judgements, legal literature, and miscellaneous parliamentary documents, including, for example parliamentary questions and minutes of the plenary sessions. The hierarchical relations of the constitution, act, decree, and administrative act are presented by their vertical relation to each other and by the line connecting them. The rectangle above the national legal documents is intended to remind the user about the significance of the related international legal documents. Similarly to the actor view, additional information about all document types can be offered to the user in separate windows. From the information page related to a document type, a hypertext link leads to a search form allowing the user to target the search to the selected document type.

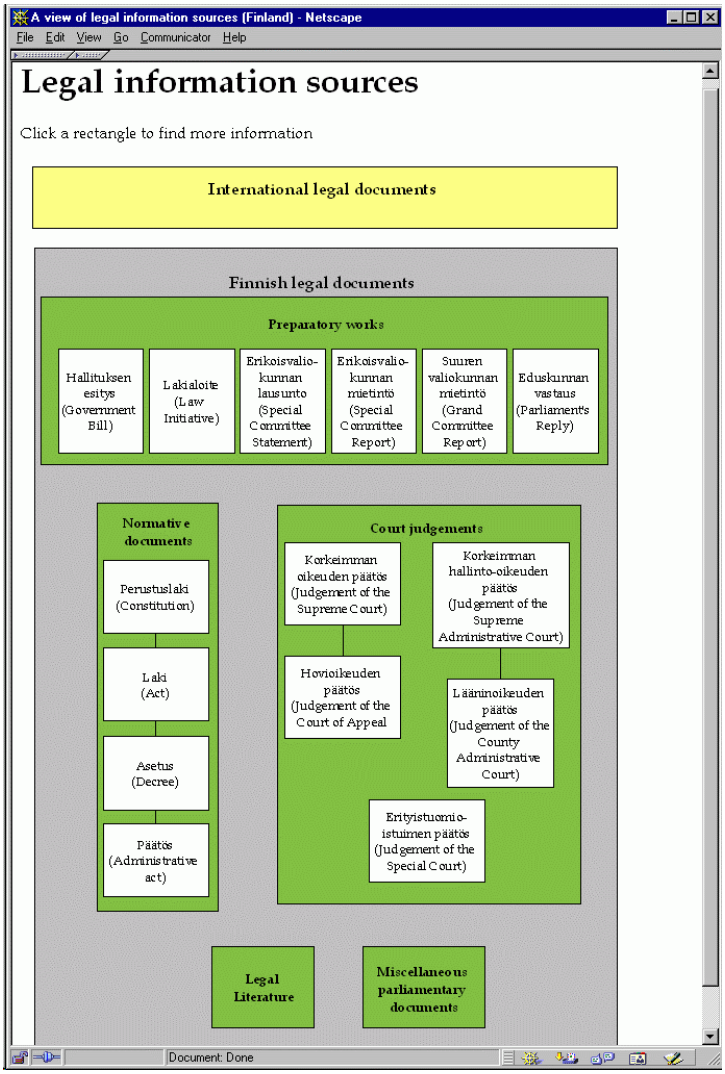


**Figure 3. Legal information source view of Finland**

**Process view**

The process view gives a graphical overview of the most significant activities related to the creation of legal sources in a certain legal system. It contains information about the order of activities, the actors responsible for performing the activities and the documents created by those activities. As an example of the process view the Finnish national legal system is presented in Figure 4.

The main activities of the modelled process are depicted by circles. Each circle shows both the name of the activity and the actor(s) performing the activity. The actors in the view are usually the same ones presented also in the actor view. The order of activities is expressed by solid arrows between the activities. Alternative activities are

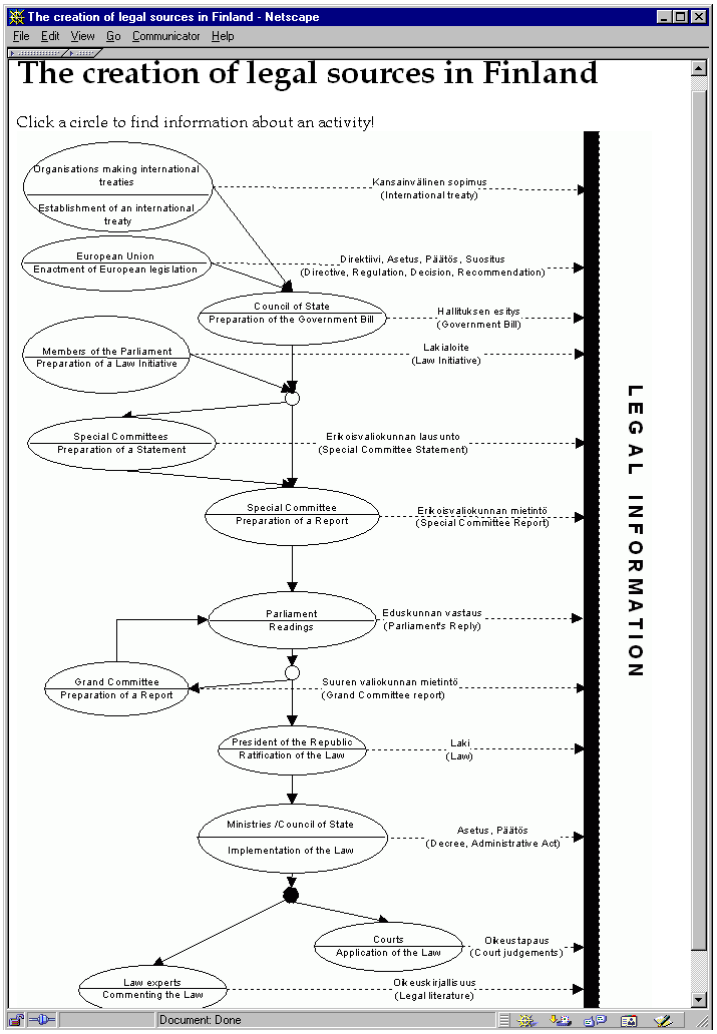


**Figure 4. Process view to the Finnish national legal system**

indicated by a hollow dot and parallel activities by a black dot. The information flow from an activity to the collective document repository is shown by a dashed arrow labelled with the type of the documents created in the activity. The document types in the process view are to a large extent the same ones presented also in the legal information source view.

Similarly to other views, additional information is linked to the activities of the process. When the users click an activity symbol in the graph, more information about the activity will appear. That information will provide a link to a search form by which the documents created by that activity can be queried.

**Implementation plans**

For the dynamic generation of graphical views the data pertaining to those views has to be formally defined. For this purpose we used the subset of SGML called XML (Extensible Markup Language), which has been developed especially for specifying document standards to be used in Web information systems (Bray, Paoli & Sperberg-McQueen, 1998). We have designed an XML document type definition by which legal systems and the different views of them can be described. The definition enables the models to be described in many languages.

## 6. DISCUSSION AND CONCLUSION

The development of the interface introduced above has been motivated by the user needs revealed in the EULEGIS project and its predecessor project RASKE. The users had lack of knowledge concerning the context of legal documents, especially in cases where the documents originated from a foreign legal system. The complexity and heterogeneity of the information sources also caused problems. The solutions developed in the EULEGIS project are intended to help the users to better understand the legal systems they are interested in, and to retrieve legal information from various information sources within those systems. The core of our approach is to collect information about legislative processes, actors involved in the processes, and document types created during the processes, and to offer the information to the users through graphical models. The models also serve as a path to the search forms by which queries to the actual legal databases can be formulated.

The complexity of the legal domain naturally causes problems also in the implementation of the suggested approach. Since legal information related to a legal system is usually scattered in different databases, implementing the solution introduced above requires negotiations with different database providers. Aiming connections to all kinds of legal documents within a legal system hardly is a realistic goal. Thus inconsistent linking has to be accepted as a partial solution. For example, links to search forms from every activity of the process view of Figure 4 cannot be established. If the search form cannot be offered, only information about the activity is shown to the user.

In order to build a model of any legal system, a deep expertise on the modelled domain is needed. Hence collaboration between the modellers and legal experts is needed. For example, in building the graphical views to the Finnish legal system in the EULEGIS project we discussed with several experts of the legal domain. During the interviews the specialists gave us valuable feedback concerning the initial graphs we had drawn earlier on the basis of literature. The opinions of the specialists differed slightly depending on their viewpoint to the legal system. Some of the specialists were officials who draft laws in ministries, while some other worked within the Parliament as legal experts or information specialists. In building the models, the modeller has to be capable to find a reasonable compromise between different opinions.

Since a model is always an abstraction and simplification of the real world phenomenon, the modeller needs to decide which details the model should include or exclude. The fact that the models are shown as user interfaces further limits the number of details that can be presented in a model. On the other hand, the model should be expressive enough to offer to the user the most valuable information. The goal for simplification sometimes conflicts with the traditional exactness on the legal area. Models as abstractions of the complex domain are never perfect.

A problem related to modelling is the grouping problem. In modelling legal information sources, it is not always clear which document types belong into a certain category. For example, in collecting information for the legal information source view of Figure 3 there was discussion about the document types that should be included in the preparatory works. A decision was made not to show all kinds of miscellaneous parliamentary documents (e.g., the minutes of the plenary session) in the group of preparatory documents but to specify a separate group for Miscellaneous parliamentary documents. There were different opinions about the position of the Parliament's Reply. A reply may contain the final wording of an act. According to most of the experts interviewed, the Parliament's Reply can be considered as a preparatory work, because the actual law will be published after the Parliament has given its reply.

In spite of the problems related to the implementation of the approach, the feedback of the end users encourages continuing the work on the area. Prototype versions of the different views have been presented to representatives of EULEGIS user groups. They considered the visualisation of legal systems as an important means for making the legal processes more understandable. A prototype implementation of the

interface together with retrieval capabilities will be developed in the EULEGIS project. Efficient implementation of these capabilities, however, requires more research. One important research area concerns the conversion of the XML specifications into the graphical form quickly and smoothly. Another area for further research is the development of a tool for guiding and assisting a modeler to model a legal system.

**Acknowledgments**

The helpful cooperation and extensive knowledge of experts in the Finnish Parliament and ministries as well as in many other organisations has been extremely valuable. The financial support of the Telematics Application Programme of the European Commission is gratefully acknowledged.

**References**

Bray, T., Paoli, J., & Sperberg-McQueen, C. M. (1998). Extensible Markup Language (XML) 1.0. W3C Recommendation 10-February-1998. http://www.w3.org/TR/REC-xml.

Gallagher, M., Laver, M., & Mair, P. (1995). *Representative Government in Modern Europe*. McGraw-Hill. New York.

Goldfarb, C.F. (1990). *The SGML Handbook*. Edited and with forward by Y. Rubinski, Oxford University Press.

Kiuru, J.(1994). Att motivera ock hänvisa inom juridiken, in *Rättslig informationssökning i databaser*. Nordisk Årsbok i Rättinformatik, Fritzes, 90 - 111.

Kiuru, J., Salminen, A., & Chen, S.-S. (1999). Digital archiving of legislative documents: A case study of Finland. *NSF Workshop on Data Archival and Information Preservation*, March 26-27, 1999, Washington DC. http://cecssrv1.cecs.missouri.edu/DA+IPpapers/finland.html

Salminen, A. (2000). Methodology for document analysis, to appear in A. Kent (Ed.), *Encyclopedia of Library and Information Science*. New York: Marcel Dekker, Inc.

Salminen, A., Kauppinen, K., & Lehtovaara, M. (1997). Towards a methodology for document analysis. *Journal of the American Society for Information Science* 48 (7), 644-655.

Salminen, A., Lehtovaara, M., & Kauppinen, K. (1996). Standardization of digital legislative documents, a case study. *Proceedings of the 29th Annual Hawaii International Conference on System Sciences*, (pp. 72-81). IEEE Computer Society Press.

Salminen, A., Lyytikäinen, V., & Tiitinen, P. (1999). Putting documents into their work context in document analysis. To be published in *Information Processing & Management*.