Fill the Gap in the Legal Knowledge Modelling

Monica Palmirani[1], Giuseppe Contissa[2], and Rossella Rubino[1]

[1] CIRSFID - University of Bologna
{monica.palmirani,rossella.rubino}@unibo.it
[2] CODEX - Stanford University
contissa@stanford.edu

Abstract. There is a gap between the legal text description in XML trends and the legal knowledge representation of the norms that from the text starts. This gap affects the effectiveness of the legal resources exploitation and the integrity of the legal knowledge on the Web. This paper presents a legal document model for managing the legal resources in integrated way and linking all the different levels of representation[1].

1 Introduction

The AI & Law community dedicated the last twenty years to model the legal norms using different logics and formalisms [14]. The methodology used starts from a re-interpretation of the legal text by a Legal Knowledge Engineer that extracts the norms, applies models and theory using a logic and finally represents them with a particularly formalism. In the last ten years several Legal XML standards were arisen for describing the legal text [9] and rules (RuleML, RIF, SWRL, etc.). In the meantime the Semantic Web, in particularly the Legal Ontology research, combined with the NLP extraction of the semantic [13], has borne a great impulse to the modelling of the legal concepts [3]. In this scenario there is the urgent need to close the gap between the text description, definitely using XML techniques, and the norms modelling in order to realise an integrated and self-contained representation of the legal resources available on the Web.

The reasons of this urgent need are primarily four.

- the legal knowledge is now presented in a disjointed way by the original text that inspired the logical modelling. This disconnection between legal document management and logic representation of the norms affects strongly the real usage of the legal resources in many applications of the law in favour of the citizens, public administrations, enterprises;
- the change management over the time of the legal document, especially the legislative one that for its nature is variable and subject to frequent modifications, significantly affects the coordination between the text and the rules that should be remodelled;

[1] This work was partially performed inside of the Estrella IST6-project n. 2004-027655. We would like to thanks Thomas Gordon for encouraging us in this vision and also for providing to the authors useful comments during several discussions on the LKIF.

G. Governatori, J. Hall, and A. Paschke (Eds.): RuleML 2009, LNCS 5858, pp. 305–314, 2009.

- the legal validity of the text as published by the official authority should be preserved by any manipulation. On the other hand it is important to connect legal document resources, that includes itself many legality values, with the multiple interpretations coming from the legal knowledge modelling;
- a theory of legal document modelling able to separate clearly the many layers of representation of the resource: content (text), structure of the text, metadata on the document, ontology on the legal concepts expressed in the text, legal content modelling (normative part of the text) is fundamental for preserving over the time the digital legal text enriched by many semantic annotations, including also logic representation of the norms.

This paper aims to present a new approach for joining two fields of the legal research, AI&Law and Legal Document Management, for realising the structure of the Semantic Web in favour to a concrete application of the legal knowledge information in the retrieval and in the legal reasoning field. A pilot case and the related methodology is presented for better explaining the model presented. Finally the conclusions comments the results and critical issues.

2 Layers of Legal Document Modelling

The state of the art of the last ten years produced plenty of Legal XML standards for describing the document as legal resource. We can divide these standards in four categories:

- the *first generation* of Legal document XML standard[2], was oriented mostly to describe the legal text and its structure with an approach near to the database entities or the typography-word processing paradigms;
- the *second generation* posed more attention to the document modelling and to the description of text, structure and metadata[3]. Nevertheless the descriptiveness of the elements was not preceded by an abstract analysis of the classes of data and the result is a very long list of tags, a complex inclusions of DTDs or XML-schema, with a frequent overlapping between metadata and text definition and a weak instruments for linking the text to any other layers;
- the *third generation* is based on pattern. The pattern defines the properties of the class and its grammar, content model, behaviour and hierarchy respect the other classes, so any additional tag belongs to an existing abstract class and in this way it is preserved the consistency over the time. A strong attention to divide the text, structure, metadata and ontology is a primary principle in order to track in robust way any new layer put on the top of the pure text. Because the pattern defines general rules that no longer impose real constraints in the mark-up action, so the clarity of design scarifies the prescriptiveness [4];
- the *four generation* uses the pattern jointly with co-constraint grammar like, among the others, RELEX NG [10], Schematron [6], DSD [8], etc. for resolving above mentioned problem of lack of prescriptiveness.

[2] Like EnAct or Formex.
[3] As NiR or Lexdania.
[4] Akoma Ntoso [15] and CEN/Metalex are examples of this approach.

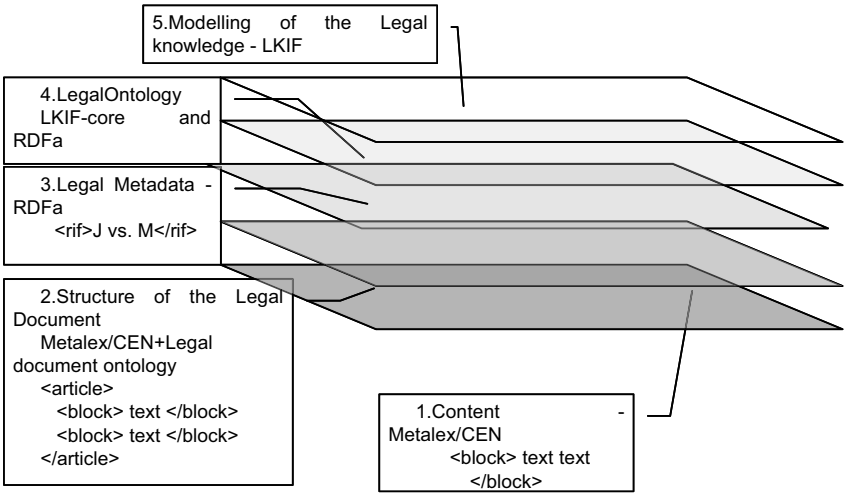

Fig. 1. Layers of representation in the Legal Document Modelling

For now we are using CEN/Metalex, as a transformation of NiR, jointly with LKIF [5] that supports RELEX NG.

The aim is to define a general legal document architecture able to describe all the following five layers (see Fig. 1) with a unique syntax or with reasonable hooks for integrating, in a cleanest way, all the different layers without confusion:

- **text:** part of the document officially approved by the authority with the legal power;
- **structure:** of the text: part of the document that states an organisation of the text;
- **metadata:** any information that was not approved by the authority in the deliberative act;
- **ontology:** any information about the reality in which the document act a role (e.g. for a judgement the juridical system concepts) or any concept called from the text that needs a modelling;
- **legal knowledge representation:** the part of the interpretation and modelling of the meaning of the text under legal perspective. Several XML standards are present in the state of the art for managing rules (RIF, RuleML, SWRL), nevertheless LKIF seems to provide a flexible language able to describe different possible theories or logic models (propositional, predicative, argumentative, non-monotonic, deontic, defeasible, etc.) more fitted for the legal domain.

3 Pilot Case Scenario

A pilot case scenario is used for explaining in concrete the Legal Document Modelling. It is based on the Italian Saving Tax Law (Legislative decree No 48 of 18 April 2005) that implements the EU Directive (Council Directive 2003/48/EC) in the same domain. One of the most important problems in the national legislative system of the EU member states is to implement, by transposition, the EU Directive taking in consideration the mandatory norms of the supra-national regulation and in the same time to be coherent with the national legislation.

The Italian pilot case aims to model the EU Directive and the national legislation concerning the Saving Tax Law with LKIF and to detect the inconsistencies between the two legislative norms using LKIF standard. On the other hand this pilot case helps the legal drafting activity and versioning over the time of the Italian corpus because the EU Directive has been modified three times until now and the Italian Saving Tax Law has consequently needed to be adapted.

For implementing the pilot case we adopted the following methodology in order to model, describe and represent the different levels of the legal knowledge information:

- **text, structure, metadata:** the EU Directive and the Italian Law were marked-up in NiR and CEN/Metalex XML standard using Norma-Editor[5] during the mark-up and validation actions;
- **ontology:** an ontology of the EU Directive and the Italian Law was made in order to model and define the macro-concepts specific for the Tax domain;
- **rules:** the EU Directive and the Italian Law were modeled in LKIF-rule syntax manually by a legal knowledge engineering. The LKIF-rule files were imported inside of an inference engine[6] properly customized by CIRSFID with a specific dialog interface for checking the consistency and for managing the comparison between the EU directive and the Italian Law.

Finally all the legal resources and the results were stored and delivered on the Web with a native XML database[7]. In this way all the legal resources (text, ontology, logic representation) were interconnected and presented using Internet interface and API for fostering the legal knowledge through information retrieval engine, reasoning engine, application layer.

4 Legal Text Description and Representation

The document texts were marked up firstly with the national Italian standard NormeinRete for grasping as much as possible the full descriptiveness of the legal document using a specific national reached mark-up. After that we translated the NIR-XML in CEN/Metalex using Norma-Editor [11] for working with a standard based on pattern and more flexible to manage the other layers linking.

Two mechanisms are present in CEN/Metalex for realizing the connection between multi-layers division above presented:

- RDFa assertions in the XML, for linking the structural part (second level) with the full ontology of classes (fourth level), with clear distinction and no overlap of relations and literal values;
- URIs naming convention, based on FRBR [1], is defined to identify the legal resources. This naming permits to link the LKIF-rules or arguments assertion directly to the part of the text involved in (connection between second level with the fifth level).

[5] Norma-Editor is a specialized legal drafting editor developed by CIRSFID and based on Microsoft-Word. It is able to convert the final mark-up action in different XML formats.

[6] RuleBurst engine by RuleBurst Europe Ltd., now a branch of Oracle Corporation UK Ltd.

[7] eXistrella native XML database presented in Jurix2008 LegalXML Workshop, Palmirani, Cervone.

A fragment of XML text in the Table 2 demonstrates the twofold mechanism: (i) `rdfa:property` says that the current metadata is-a member of the `lkif-document:URI` ontology class; (ii) the `rdfa:content` value specifies the URI of the resources that will be used in the LKIF XML file.

Table 1. Fragment of CEN/Metalex Saving Tax Law: identification of the document

```
<meta xmlns:rdfa="http://www.w3.org/TR/xhtml-rdfa-primer" .
    name="manifestationURI"
    id="metalex_d1e5"
    value="it/act/2005-04-18/84/2005_04_18_dlgs_84.xml"
    rdfa:property="[lkif-document:URI]"
    rdfa:content="/it/act/2005-04-18/84/ita@/2005_04_18_dlgs_84.xml"/>
```

5 LKIF Ontology of the Italian Savings Tax Law

The aim of the ontological level is to model any concepts (definition and properties) that is useful for completing the understanding of the rules or arguments represented in LKIF. Secondary the OWL syntax permits to make some consistency check on the concepts using some existing engine like Pellet. The rationale is to define in the ontology any static concepts (T-Box) of the context useful for the understanding of the rule and arguments. Some classes of this ontology[8] are presented hereinafter:

- ***income.*** According to the LKIF ontology the *income* can be classified as a `top:Mental_Concept`. We consider, here, only the subcategory of ***savings income*** in the form of interest payments. The form of an income has been modelled as the property *form* of the class *savings income.*

Fig. 2. Savings_income class

- Every ***payment***, which is an `action:Action` in the LKIF ontology, has a subject, a payer, a beneficiary, the amount of the payment, the medium adopted to make the payment and finally the date and the place where the payment has been performed. An ***interest payment*** is a payment whose object is an interest.

Fig. 3. Interest_payment class

[8] They are an extension to the LKIF-core ontology [4].

- The ***beneficial owner*** is the individual, that is a `j.1:Natural_Person` in the LKIF ontology, who receives an interest payment or for whom an interest payment is secured. We defined some properties which are useful: the tax identification number, the place and the date of birth, the name of the beneficial owner, his address and his country.

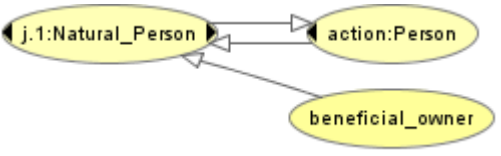

Fig. 4. Beneficiary_owner class

- The ***paying agent*** is any economic operator who makes an interest payment to the beneficial owner or secures a savings income payment for the beneficial owner. The ***economic operator*** is the individual or body which actively takes part - from the side of demand and of offer - in a specific market.

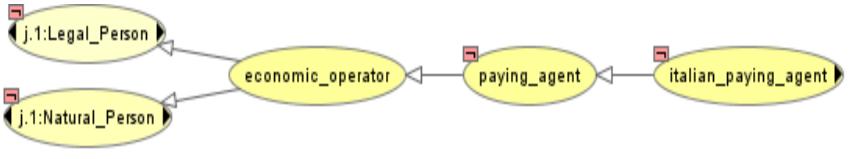

Fig. 5. Paying_agent class

These OWL classes are linked both in the CEN/Metalex XML file and in the LKIF-rules modelling using two different syntaxes base on the same principles in order to maintain consistent and self-contained the legal document representation in the Web:

- inside of the CEN/Metalex XML we focused our attention on linking the fragment of the text with the ontology class using RDFa assertion by the `rdfa:rel` and `rdfa:href` attributes.

Table 2. Fragment of CEN/Metalex Saving Tax Law: assertion in RDFa using the class interest_payment

```
<htitle name="rubrica" id="metalex_title_art2 " rdfa:about="" rdfa:rel="rdf:type"
rdfa:href="[savings:interest_payment]">Interest payments.</htitle>
```

- inside of the LKIF-rule we specify the `pred` attribute's value with the name of the predicate that is defined inside of the ontology class `saving:paying_agent`. This mechanism permits to use classes, properties and definitions uniquely modelled inside of the ontology with a great benefit for the consistency and interoperability.

Table 3. Fragment of LKIF Saving Tax Law modelling: paying_agent predicate

<s pred="savings:paying_agent"><v>X</v> shall communicate the information concerning the payment to Revenue Agency under 1-1b</s>

6 LKIF Modelling

The main goal was to represent in LKIF the rules for fulfilling the requirements of the Italian Pilot, which involves a comparison of EU and Italian norms in a legal drafting perspective. A high degree of granularity was requested and the knowledge representation should have been isomorphic to the maximum possible extent, as described by Bench-Capon [2] and Karpf [7]. Each legal source is represented separately, preserving its structure and the traditional mutual relations, references and connections with other legal sources, thanks to the fact that structural elements in the source texts correspond to specific elements in the representation. The representation of the legal sources and their mutual relations should also be separated from all other parts of the model, in particular the representation of queries and facts management.

Therefore, each norm from the legal source text was represented in a correspondent rule (or set of rules). No deviations were made, as long it was possible, from the original structure of the text, even when it was redundant or confusing.

The representation of the Italian Savings Tax Law was carried out as follows: the law was analyzed to find the core part to be modelled, keeping a correspondence with the modelled part of the Directive. The following excerpt is a fragment representation of Art. 1 comma 1, first paragraph[9].

The rules are represented with `head` and `body`, the `body` is a set of sentences (<s>) concatenated with Boolean operators. In case of `exception` we should specify a new rule with `pred` value "execption": this produce a duplication of the rule and a verbose representation and a not efficient method for expressing the exception of exception.

The `<sources>` block specifies the URIs where to find the corresponding part of the rule into the text with a granularity since the word.

[9] *"Art. 1. Subjects held to the communication - 1. Banks, investment firms, Poste italiane S.p.a., security investment fund management companies, financial companies and trust companies, resident in Italy, shall communicate to the Revenue Agency the information concerning the paid interest or the interest for which the immediate payment is secured for the immediate benefit of individuals, that are beneficial owners, resident in another Member State; for this purpose the individuals shall be considered beneficial owners of interest if they received the payment as final recipient."*

Table 4. Fragment of LKIF-XML modelling rules in Italian Saving Tax Law

```
<?xml version="1.0" encoding="utf-8"?>
<?oxygen RNGSchema="LKIF2.rnc" type="compact"?>
<?xml-stylesheet type="text/css" href="LKIF2.css"?>
<lkif xmlns:savings="savingsTax_ita.owl">
  <sources>
    <source element="#ita_savings"      uri="/it/act/2005-04-18/84/eng@/main.xml"/>
    <source element="#s1_ita"           uri="/it/act/2005-04-18/84/eng@/main.xml#art1"/>
    <source element="#s1-1a-ita"        uri="/it/act/2005-04-18/84/eng@/main.xml#art1-com1"/>
    <source element="#s1-1a_01-ita"     uri="/it/act/2005-04-18/84/eng@/main.xml#art1-com1"/>
    <source element="#s1-1a_02-ita"     uri="/it/act/2005-04-18/84/eng@/main.xml#art1-com1"/>
    <source element="#s1-1a_03-ita"     uri="/it/act/2005-04-18/84/eng@/main.xml#art1-com1"/>
  </sources>
  <theory id="ita_savings">
    <rules>
      <rule id="s1-1a-ita">
        <head>
          <s pred="savings:paying_agent"><v>X</v> shall communicate the information concerning
the payment to Revenue Agency under 1-1a</s>
        </head>
        <body>
          <and>
            <s><v>X</v> is a subject listed in 1-1</s>
            <s><v>X</v> is resident in Italy</s>
            <s><v>X</v> pays or secures interest to <v>Y</v></s>
            <s><v>Y</v> is an individual</s>
            <s><v>Y</v> is a beneficial owner</s>
            <s><v>Y</v> is resident in another member state</s>
            <s><v>X</v> operates as debtor or charged to pay</s>
          </and>
        </body>
      </rule>
      <rule id="s1-1a_01-ita">
        <head>
          <s><v>X</v> is a subject listed in 1-1</s>
        </head>
        <body>
          <or>
            <s><v>X</v> is a bank</s>
            <s><v>X</v> is an investment firm</s>
            <s><v>X</v> is poste italiane spa</s>
            <s><v>X</v> is a security investment fund management company</s>
            <s><v>X</v> is a financial company</s>
            <s><v>X</v> is a trust company</s>
          </or>
        </body>
      </rule>
      <rule id="s1-1a_02-ita">
        <head>
          <s><v>X</v> pays or secures interest to <v>Y</v></s>
        </head>
        <body>
          <or>
            <s><v>X</v> pays interest to <v>Y</v></s>
            <s><v>X</v> secures interest payment for immediate benefit of<v>Y</v></s>
          </or>
        </body>
      </rule>
```

Table 5. Fragment of LKIF-XML modelling rules in EU Saving Tax Directive

```
<rule id="par-2-1-exception-b2-EUSD">
 <head>
  <s pred="rule:excluded"> <c>par-2-pa1-EUSD </c>excluded<s>beneficial-owner<v>x</v>
   </s></s>
  <s  pred="rule:excluded"> <c>par-2-pa2-EUSD</c> excluded<s>beneficial-owner<v>x</v>
   </s></s>
 </head>
 <body>
  <and>
   <s>individual<v>x</v>
   </s>
   <s>provides-evidence<v>x</v>
    <s>entity-taxed-on-its-profits<v>y</v>
    </s></s>
   <s>provides-evidence<v>x</v>
    <s>acts-on-behalf-of<v>x</v><v>y</v>
    </s></s>
  </and>
 </body>
</rule>
```

7 Conclusions

The pilot case underlines several considerations concerning the gap between the five levels of information that we need to model for describing a legal document:

- the granularity of the XML document marked-up is not isomorphic to the rules and statements modeling in term of part of text (word, paragraphs, etc.);
- the relationship between rules and text is a N:M cardinality, so both the LKIF-rules and the CEN/Metalex need to improve their syntax for implementing a smart mechanism able to capture, without duplication, the multiple referencing;
- the interaction between rules in LKIF and concepts expressed in the ontology, since the definitions can be considered too complex to be inserted into the ontology, also because some of them are rules by themselves;
- the relationship between the XML document and the ontology it is useful even if it is not in contrast with the usage of the same classes in the LKIF-rules;
- the legal document change over the time as well as the ontology definition so in the LKIF-rule it is necessary a mechanism for managing the dynamicity over the time as well as into ontology;
- the non-monotonic dimension of the law, the exception of exception convinced us to include some extension in LKIF for implementing the defeasible logic paradigm strictly linked with the text.

From these conclusions we learnt the lesson that we need a strong architecture of the legal document divided in five levels. For permitting a coordination between these layers we need standards designed with mechanisms for interconnecting them in asynchronous way. Our future work will go in the direction to fill the gap between the different markup granularity and temporality among the different layers. Finally we aim to implement the defeasible logic extending the LKIF schema.

References

1. Bekiari, C., Doerr, M., Le Boeuf, P.: International Working Group on FRBR and CIDOC CRM Harmonization. FRBR object-oriented definition and mapping to FRBRER (v. 0.9 draft) (2008), http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/FRBR_oo_V0.9.pdf (accessed August 20 2009)
2. Bench-Capon, T., Coenen, F.: Isomorphism and legal knowledge based systems. Artificial Intelligence and Law 1(1), 65–86 (1992)
3. Boer, A., Radboud, W., Vitali, F.: MetaLex XML and the Legal Knowledge Interchange Format. In: Casanovas, P., Sartor, G., Casellas, N., Rubino, R. (eds.) Computable Models of the Law. LNCS (LNAI), vol. 4884, pp. 21–41. Springer, Heidelberg (2008)
4. Breuker, J., Boer, A., Hoekstra, R., Van Den Berg, C.: Developing Content for LKIF: Ontologies and Framework for Legal Reasoning. In: Legal Knowledge and Information Systems, JURIX 2006, pp. 41–50. ISO Press, Amsterdam (2006)
5. Gordon Thomas, F.: Constructing Legal Arguments with Rules in the Legal Knowledge Interchange Format (LKIF). In: Computable Models of the Law, Languages, Dialogues, Games, Ontologies 2008, pp. 162–184 (2008)
6. Jelliffe, R.: The Schematron Assertion Language 1.5, http://www.ascc.net/xml/resource/schematron/Schematron2000 (accessed June 20, 2009)
7. Karpf, J.: Quality assurance of Legal Expert Systems, Jurimatics No 8, Copenhagen Business School (1989)
8. Klarlund, N., Møller, A., Schwartzbach, M.I.: DSD: A Schema Language for XML. In: Proceedings of the third workshop on Formal methods in software practice, Portland (2000)
9. Lupo, C., Vitali, F., Francesconi, E., Palmirani, M., Winkels, R., de Maat, E., Boer, A., Mascellani, P.: General xml format(s) for legal sources - Estrella European Project IST-2004-027655. Deliverable 3.1, Faculty of Law. University of Amsterdam, Amsterdam (2007)
10. Murata, M.: RELAX (REgular LAnguage description for XML) (2000), http://www.xml.gr.jp/relax (accessed June 20, 2009)
11. Palmirani, M., Brighi, R.: An XML Editor for Legal Information Management. In: Proceeding of the DEXA 2003, Workshop su E-Government, Praga, September 1-5, pp. 421–429. Springer, Heidelberg (2003)
12. Peruginelli, G., Ragona, M. (eds.): Law via the Internet Free Access, Quality of Information, Effectiveness of Rights, EPAP, p. 494 (2009)
13. Proceeding of the 12th International Conference on Artificial Intelligence and Law, Barcelona, June 8-12, p. 243. ACM, New York (2009)
14. Sartor, G.: Legal Reasoning: A Cognitive Approach to the Law. In: Treatise on Legal Philosophy and General Jurisprudence, vol. 5. Springer, Berlin (2005)
15. Vitali, F.: Akoma Ntoso Release Notes, http://www.akomantoso.org (accessed August 20, 2009)