Legislation in the semantic web

Giovanni Sartor

Law department, European University Institute, Florence
Cirsfid, University of Bologna
giovanni.sartor@eui.eu

3 novembre 2007
Outline

1. Laws in the web
2. The legal semantic web
Preserving the role of legislation in the information age

- align legislation to its purposes: anticipate and monitor its impact on the legal systems and on society
- promote the critical analysis: support making legislative proposals and rationally debating their merits
- align legislation to citizens’ reasoned preference and views: enable citizens to express their view and preferences, provide them with feedback about choices of their representatives
- enable citizens to anticipate and control the application of laws: provide access laws and cases, ensure understandability
- coordinate legislation with other legal sources (national, foreign, international): take into account other legal sources, learn from others, develop joint initiatives
Preserving the role of legislation in the information age: ICT contribution

- align legislation to its purposes: ICT can provide information on the legal and social effects of legislative acts
- promote the critical analysis: ICT can support communication and reasoned debate within the legislative process
- align legislation to citizens’ reasoned preference and views: ICT can support communication and democratic debate
- enable citizens to anticipate and control the application of laws: ICT can provide tailored and updated information about law in force and its application
- coordinate legislation with other legal sources (national, foreign, international): ICT can support information, support communication, coordination and cooperation
Some tasks for a parliamentary information systems (beside administration)

Computer support for
- preparing documents (drafting tools, linguistic tools, amendments management, etc.)
- accessing information (information retrieval, legal and factual databases)
- simulation (“What if analyses”) of legal and social impacts
- workflow management (lifecycle, security, timely involvement of relevant individuals)
- citizens’ access to the law (multi-channel publication, on paper, web, and thorough third parties, law in force),
- political dialogue (e-petition, e-consultation, discussion fora)
- institutional dialogue (interaction with courts, administration, other parliaments)

Not only MPs are interested: ICT enables openness to all
Parliamentary information systems
Functions of Parliamentary IS

- MPs
- Experts
- Public
- NGOs
- Courts
- Drafters
- Lobbyist
- Committees

- Political Dialogue
- Workflow Management
- Drafting
- Concept Management
- Amendment Versioning
- E-Petitions
- E-Consultation
- Electronic Texts and Data
- Broadcasting
- Impact Evaluation
- Retrieval
- Public Administration
- Electronic Texts and Data

Giovanni Sartor

Legislation in the semantic web
A huge amount of legislative information is nowadays available on the web:

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

We would like to:

- Automatically obtain relevant textual information, opportunely selected and integrated
- Automatically process information embedded in texts (e.g. generate the law in force, control deadlines, apply rules, etc.)
Two strategies:

1. automatically understand and process natural language texts. Promising but limited results (for instance in automatic translation and classification)
2. insert in the web (within natural language texts or elsewhere) information that is machine understandable and can be process with the available advanced techniques (W3C approach)

The second strategies has produced some results:

- better accessibility of textual information;
- supporting automatic integration of textual information;
- overcoming distinction between textual information and structured data (all legal information on the web, both texts and data, and various combinations of them)
Processable legal semantics?

- **Semantic web**: information partly machine understandable, which can be automatically processed according to its meaning
- **Legal semantic web**: legal (legislative) information partly machine understandable, automatically processable according to its legal meaning
Tim Berners-Lee’s vision

- Trust
- Proof
- Logic
- Ontology vocabulary
- RDF + rdfschema
- XML + NS + xsmlschema
- Unicode
- URI (identificatori)
- Data
- Rules
- Self. desc. doc

Legislation in the semantic web
Tim Berners-Lee’s vision mapped to the law

- Logic (norms and facts)
- Ontology vocabulary (legal concepts)
- RDF + rdfschema (description of legal resources)
- XML + NS + xsmschema (standards for legal documents)
- Unicode
- URI (identifiers for legal resources)
- Trust (certainty)
- Proof (inferences, arguments)
- Data
- Data
- Data
- Self. desc. doc
The legal semantic web consists of automatically processable information (meanings) concerning:

- legal documents
- the law itself
  - the law contained in the document (which speaks the law, stating norms and defining concepts)
  - legal norms (without a precise textual reference)
  - legal facts (the facts to which legal norms apply)
Machine processable legal contents:

- structures of legal texts
- references between legal texts
- the life-cycle of legal texts
- modifications of legal texts
- semantic of legal terms (thesauri, computational lexicons, light ontologies)
- legal concepts and their definitions (formal ontologies)
- semantical structures of normative language (semantic annotation of legal texts)
- norms (logical formalisation of normative contents)

If we add machine processable information to texts, then ICT can help us in preparing, storing, communicating, using such texts.
The double vision of the law

legal semantic web

ontologies

norms

texts

metadata

inferences/arguments

social and legal reality

Giovanni Sartor

Legislation in the semantic web
Perspectives:

- the web can provide all with accessible legal information, tailored to their needs
- the web can be a space of democracy, a forum for political and legal debate.
- the web can support communication between legal agents

For this purpose we need to build a legal semantic web but:

- it must be accessible to everybody
- it must be accessibly to all systems and devices
- it must empower everybody as a producer of legal information

Legislation is a fundamental part of that information, and Parliaments can have a leading role in creating it.
The legal semantic web represents an opportunity for the legislation

- maintenance of legal sources
- improvement of legal del drafting
- legislation based upon knowledge and dialogue
- publicity of procedures and information
- dialogue between sub-national, national, and international institutions

Compliance with (reasonable and) shared standards is a precondition for this opportunity to be realised.
Standards based on XML (eXtended Markup language), including:

- **Markup**: computer-processable information added to a text: `<title>Privacyact</title>`
- **URI**: universal resource identifiers: `stato:legge:1997-07-16;254@2000-12-03`
- **Metadata**: information about the document, which is included in the document (e.g. the lifecycle)

All of these elements must be provided according to an open, non-proprietary model.
Standards about what?

We need to specify in a precise way, understandable to a machine, and shared by all users, the following elements:

- the identification of the legal source
- the organisation of its content (structure)
- links to other sources
- textual modifications
- information about the source (metadata)
- concepts
- norms

NB: we need an incremental approach: each level is a precondition of the other, but no need to go beyond what is required and feasible.

NB: each addition of machine processable information adds to what can be done with the text, to the support we can get from ICT tools.
What standard-based tools

- editors
- converters
- name resolvers
- validators
- document management systems (with search engines, hypertext functionalities, XSLT support and versioning facilities)
- retrieval tools
- workflow managers
- publishing systems
- knowledge-based systems
- etc.
Why standards

Because they enable

- preservation
- communication
- processing
- openness
- enrichment
- control
- decentralisation
- subsidiarity
- investment
- sharing
- competition
- cooperation
- neutrality
- progress
XML-based standards for legislation

- National standards, in many countries (e.g. Norme in rete, the Italian standard for legislative documents and regulations)
- Emerging international interchange standards: Metalex, developed by the Metalex-CEN consortium
- Emerging international standards: Akoma-Ntoso
A centralised standard-based infrastructure for legal (from Jon Bing)
A decentralised standard-based infrastructure for legal (from Caterina Lupo)
A decentralised standard-based infrastructure for legal (from Caterina Lupo)

Future scenario

Norme in Rete (or new institutional portal)

LEGISLATION IN THE SEMANTIC WEB

Giovanni Sartor
Legislation in the semantic web
Legal texts enriched with machine processable information, which are
- produced during the legislative process
- containing information which contributes to direct the workflow
- preserved to record the workflow
- transformed into new documents as the workflow proceed (transferring the both the text metainformation)
- made accessible to everybody (without additional processing)
- open to everybody inspection
- distributed on a non-discriminatory basis

NB: we need an incremental approach: each level is a precondition of the other, but no need to go beyond what is required and feasible
Giovanni Sartor
giovanni.sartor@eui.eu
Semantic Standards

Maria Angela Biasiotti
Semantic Metadata Role and Scenario

• Semantic Resources or metadata
  Definition
  Typologies
  • State of the art in existing semantic resources
  • Future Approach.
  - Integrating Semantic Resources
  - Mapping, merging lexical resources
  - Aligning lexical resources and ontologies
Migliorare l'accesso all'informazione

La semantica indipendente da sintassi e da contesto (pragmatica) consente di catturare i contenuti svincolati dalle struttura linguistica. Perciò:

- la ricerca full-ext e su documenti non strutturati (pagine web) è guidata dai concetti non dalle parole
- gli utenti non esperti possono utilizzare termini del linguaggio comune per accedere a conoscenza tecnica
- si possono rintracciare informazioni espresse in lingue diverse

Daniela Tiscornia
Semantic Metadata Role and Scenario

Semantic 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.
# e-Government Metadata Standard

**Version 3.0  29 April 2004**


## 2.23 Subject

<table>
<thead>
<tr>
<th>Definition</th>
<th>A topic of the content of the resource.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obligation</td>
<td>Mandatory (Category refinement).</td>
</tr>
<tr>
<td>Purpose</td>
<td>Enables the user to search by the topic of the resource.</td>
</tr>
<tr>
<td>Notes</td>
<td>Two different approaches to information seeking are commonly used: browsing through a directory (also known as drill-down or navigation) and searching by direct entry of keywords. The Category refinement is intended to support the first of these, browsing through a directory of broad classes, while the Keyword refinement supports direct search. The values for all Subject refinements should be drawn from encoding schemes, also known as controlled vocabularies, thesauri or authority lists. There are different encoding schemes for each refinement. It is important to tag each value entered to indicate the source scheme.</td>
</tr>
</tbody>
</table>
In a wide meaning, which includes both ontologies and computational lexicon, Semantic Resources provide a "shared vocabulary” for Semantic Webtools, aimed at supporting:

- information integration, semantic interoperability,
- knowledge sharing, building machine understandable knowledge bases, etc.

More specifically, within a Legal Standard Framework, they represent:

- sources of metadata for content description
- allowing semantic annotation.
Semantic Resources Role and Scenario

- Web = documents (before reserved to the access and knowledge of experts) are now more and more accessible and available
- From text-based web to the semantic web where information are processed by their meanings
- Semantic Web and Legal Semantic Web (processed according to the legal meaning)
- EU Commission no. 1049/2001 trovare definizione precisa
- Key-words attached to document convey all semantic information from the ontology, allowing either conceptual, cross-lingual or integrated Information Retrieval
Dal Text-based web al Semantic Web: Documento vs. Conoscenza

*Goal*

To transform text based documents into information and knowledge sources for:

- migliorare la ricerca delle informazioni
- rendere i contenuti giuridici comprensibili ai cittadini
- superare le barriere linguistiche
- rendere i contenuti di Internet comprensibili ai programmi (agenti)
- consentire l'automazione di schemi di ragionamento giuridico

Daniela Tiscornia
Legal Semantic web

The legal semantic web consists of automatically processable information (meanings) concerning:

- legal documents
- the law itself
- the law contained in the document (which speaks the law, stating norms and defining concepts)
- legal norms (without a precise textual reference)
- legal facts (the facts to which legal norms apply)

Sartor, LeX summer School
Semantic Resources Typologies

- Traditional Lexical Resources
- Semantic Nets
- Formal Ontologies

- Formalization Standards: XML, RDF, OWL
Typologies: Traditional Lexical Resources

- Controlled vocabularies (taxonomies, headings, directories, thesauri, key-words) are lists of terms organized in hierarchies (BT/NT) and linked by generic RT relations.
- No semantic constraints
- Information Retrieval oriented
Typologies: Semantic Web oriented Resources

- Semantic nets (lexical or lightweight ontologies): nets of concepts structured according to lexical, taxonomic and conceptual relations
  - Constraints over relations are based on POS (Part Of Speech) categories.
  - Concept density

Daniela Tiscornia
Metadata

- Data about data
- Structural data about data
- Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics (DESIRE Project)
Semantic Nets or Lightweight Ontologies

Simple taxonomic structures of primitive or composite terms together with associated definitions. These are the so-called lightweight ontologies, 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 community.
Ontologies

“The metaphysical study of the nature of being and existence”.

Artificial Intelligence “a shared and common understanding of some domain that can be communicated between people and application systems” [Gruber].

a formal specification of conceptualization providing the elements with which the knowledge will be expressed and not the knowledge itself thus an ontology defines the terms used to describe and represent an area of knowledge [Gruber T.R, 1994];

• a way for defining the terms used to describe and represent situations in the world [Breuker J. 2001];

• an artifact designed to account for the commitment of a language to certain conceptualization, [Guarino N., 1999].

Generally speaking an ontology defines concepts and relations comprising the vocabulary of a topic area as well as the rules for combining concepts and relations to define extensions to the vocabulary [Neches, 1991]. More specifically we can declare that an ontology specifies how to view the world, which terms we should speak about the world, which are the essential concepts expressing that world.
Ontologies Levels
(from D. Tisconia)

Vertical typology:

- Foundational Ontology
- Domain Core Ontology
- Domain Specific Ontology

Horizontal typology:

- Philosophical ontology
- Information System ontology
- Linguistic ontology
Thesaurus

Structured list of terms in a specific domain

- **Descriptors**
- **Non-descriptors**
- **Semantic relationships**
  - **Equivalence**
  - **Hierarchical (BT/NT)**
  - **Associative (RT)**
Semantic Resources implemented in Legal Information Access

• **Eurovoc, 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.

• **IATE (Inter-Active Terminology for Europe)**, the EU inter-institutional terminology database generated 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)

• **LOIS Project 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).

• **The Syllabus Project 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.
Some Existing Semantic nets

*WordNet*, 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.

*FrameNet*, online lexical resource for English, based on the principles of Frame Semantics, 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). 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.

*CYC*, 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.
Some existing Ontologies

Foundational: DOLCE, SUMO
Conclusions