

# LEGALURN: A FRAMEWORK FOR ORGANIZING AND SURFING LEGAL DOCUMENTS ON THE WEB

Caterina Lupo,<sup>1</sup> Luca De Santis,<sup>1</sup> and Carlo Batini<sup>2</sup>

<sup>1</sup> *Centro nazionale per l'informatica nella pubblica amministrazione (CNIPA)*

*Via Isonzo 21/b, I-00198 Roma, Italy*

{lupo,luca.desantis}@cnipa.it

<sup>2</sup> *Università di Milano "Bicocca"*

*Via Bicocca degli Arcimboldi 9, I-20216 Milano, Italy*

batini@disco.unimib.it

**Abstract** Identifying resources is a critical issue in the wide web information space. Several identification systems have been defined, each tailored to a specific domain or application field, and characterized by many limitations. In this paper we describe an identification system compliant to URN specification that has been defined and implemented specifically for the legal domain, while providing several innovative features. The system allows to easily manage references to juridical documents and to automate the distributed hyperlinking building process. Moreover, the system provides a resolution service associating to the logical identifier a physical resource (e.g. URL) and other facilities to ensure semantic coherence and unambiguousness in the uniform names attribution task. Finally, we briefly outline future work concerning the opportunity to investigate other relevant properties in the legal domain by representing laws as a directed graph.

**Keywords:** eGovernment, uniform resource name, identification system, legal documents

## 1. Introduction

The web information space consists of a huge quantity of documents, that is increasing very fast. Public administrations are among the most important sources of documents, even more nowadays as extensive e-government programmes are producing results.

In the eEurope action plan (Lisbona 2002) legislative documents have been classified as essential public information, i.e. their availability fulfills citizens' fundamental rights. Therefore several e-government initiatives are committed to provide free access to legislative documents through the Internet. Searching

legislative documents in the web is a difficult task, due to the complexity of documents and the variety of data sources together with the lack of standardization. Moreover, laws contain references to other laws and often introduce modification into some part of other existing norms. This strong correlation among documents, even if issued by different authorities, requires the availability of robust hyperlinking mechanism in order to access the entire useful information. From this point of view, a legislative corpus can be seen as a net, quite similar to the web.

In this paper we describe a legislative documents identification system aimed to simplify searching and to enable automatic hyperlinking among laws references. The system has been developed within the *Norme in Rete* (norms in the net) project, funded and coordinated by the National Center information technologies in the Public administration (CNIPA).

The rules adopted to build the laws identifiers are compliant with IETF Uniform Resource Name (URN) standard. The identifier, named *legalURN*, allows to automate the laws hypertext building process heavily reducing manual activities in detecting and marking-up references. Users can navigate through the references automatically detected by a specialized parser: a specific infrastructure is responsible to resolve URNs into physical resources, e.g. a web pages. Today, the *legalURN* is an Italian national standard issued as technical norm by CNIPA.

The paper is organized as follows. In Section 2 we describe the *Norme in Rete* project. In Section 3 we report a brief description of what an identifier is expected to be, whereas in Section 4 we illustrate some of the most used digital resources identification systems. Moreover, in Section 4 we provide the motivations that have driven to the *legalURN* standard definition and adoption. In Section 5 and Section 6 we illustrate the *legalURN* syntax and semantics. In Section 7 we describe the system architecture developed to support *legalURN* and the related services available to use it effectively. Finally, in the Conclusions we draw some research lines and issues that are currently being investigated.

## 2. The *Norme in Rete* project

Information technology is changing the way legal documents are managed and accessed. A growing number of initiatives are today in progress reaching significant results.

Public administrations and private organizations provide access to juridical documents through the web. All this cause a proliferation of sites, each with a partial set of documents. Standardization can give the opportunity to achieve interoperability among different sites.

Many standards and system enabling and simplifying juridical documents management process have been proposed (Palmirani and Brighi, 2002; Grandi et al., 2003; Boer et al., 2002a; Boer et al., 2002b). However, none, in the authors' knowledge, addresses the issues related to the identification of juridical documents as logical entities, that can have multiple editions with different value-added metadata or ensured quality standard. The NormeInRete project (Spinosa, 2001; Lupo and Batini, 2003) started in 1999 was proposed by the Italian Ministry of Justice with the aim of building a distributed cooperative system to access juridical documentation. Cooperation has been achieved through standardization. The project involves the major Italian computer science and law centers, such as ITTIG<sup>1</sup> and Cirsfid<sup>2</sup>, and a number of Public administrations (e.g. Chamber of Deputies, Senate of the Republic, Italian Supreme Court of Cassation, Regions and municipalities, Authorities, several Ministries, etc.). The project has achieved the following results:

- A site providing a unique access point for searching the Italian legislative corpus. The site ([www.normeinrete.it](http://www.normeinrete.it)) offers search and retrieval services operating on all Italian laws since 1948 and utilities for automated hyperlinking. The entire project documentations and other information related to the project are also available (in Italian). The site includes also 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. The site has more than 4000 visitors and about 7000 accesses to the search functions daily.
- a standard for norms XML representation. DTDs (Document Type Definition) for Italian legislation have been defined, able to represent metadata and all the significant information useful to automate legislative documents' life-cycle management. The availability of documents with XML mark-up according to shared DTDs definitions allows to provide advanced search and retrieval functions operating on heterogeneous data bases effectively.
- A standard for persistent identification of norms. This paper is focused on this standard.
- An infrastructure for identifier resolution and management. Currently, the mechanism to resolve identifier is centralized and provided within the site itself. For example, to resolve a generic legalURN *genericURN* it is sufficient to access to the url: <http://www.nir.it/cgi-bin/N2Ln?genericURN>. The resolution engine index manages more than 50.000 legalURNs.

### **3. Background on identifiers**

In this section we briefly describe some of the main features usually required for identifiers. An identifier is a sequence of symbols that can be associated to a logical or physical object unambiguously in order to distinguish it from other objects belonging to the same category: it is able to unambiguously identify a resource, when defined and used in a specific context, i.e. a namespace. Unambiguousness is a required property for an identifier. Depending on the application context for which the identifier has to be defined other properties can be required, such as uniqueness, persistence, actionability, interoperability and intelligibility. In the following paragraphs there is a brief description of these features, checked against the Uniform Resource Locator (URL), that is one of the most widely used identification system.

The uniqueness property ensures that a resource, or a class of resources, is represented by only one identifier through a one-to-one relationship. Several applications require unique identifiers: US S.S.N. or Italian fiscal code are examples of identifiers that have to be unique in their namespace. URL does not guarantee uniqueness, for it is possible to define alias; for example, the URL <http://www.normeinrete.it> and <http://www.nir.it> point to, i.e identify, the same IP address.

An identifier is persistent if it maintains the linkage to a resource even though the physical resource has been moved or some of its features have changed. For example, an URL can be used as a web page identifier, but it does not guarantee persistence, since web pages are often moved or deleted. Note that technologies can only support the persistence of identifiers and cannot ensure it.

Actionability means that the identifier can be used with a specific applicative infrastructure to obtain some result. In the simplest case, the result will be the resource bound to the identifier. URL is an actionable identifier when used with a browser; the browser exploits the domain name system (DNS) to translate it into a physical address and the HTTP protocol to obtain the resource.

Interoperability is the ability to maintain significance when the identifier is used outside the original namespace. There are different levels of interoperability: the basic one consists of the preservation of the identifier's unambiguously in the new namespace. An URL does not guarantee interoperability.

Any dumb sequence of symbols assuring unambiguousness in a specific namespace can be used to identify a resource. When the sequence is chosen in such a way that the resulting string, or any of its substrings, has a meaning related to the resource, then it is called intelligent identifier. Analyzing the string, it is possible to extract some information about the resource. URL is an intelligent identifier composed by substrings representing the protocol used to obtain the resource, the domain of the resource and its path. A detailed

description of the identifier properties reported above can be found in Paskin, 2003.

#### **4. Related identification systems**

The identification of digital resources is a problem that has been growing in interest in the last few years. Several standards, proposed for general purpose or for domain specific application, are today widely used.

The system we describe in this paper is based on the URN specification (Masinter and Sollins, 1994). URN has been initially proposed to provide a persistent alternative to the URL to identify Internet resources. The original proposal has been developed and defined more in detail afterwards. RFC 2141 (Moats, 1997) specifies the URN syntax, while guidelines for URN resolution systems, that make URN actionable, are illustrated in the RFC 2276 (Sollins, 1998). A deeper description of URN specification and its connection with URI is provided in van der Werf-Davelaar, 1999.

The Digital Object Identifier (DOI) (see DOI, 2004) is an identification system recently emerged. It allows effective identification of digital resources, providing the infrastructure needed to handle additional information related to the resources, i.e metadata, and to ensure persistence. The system is managed by the International DOI Foundation, an open membership consortium including both commercial and non-commercial partners, and has recently been accepted for standardization within ISO.

The International Standard Serial Number (ISSN ISSN, 2004) is a free of charge dumb string of 8 numerical digits used to identify periodical publications. It is managed by 77 National Centers, with a coordination center based in Paris. In order to obtain a ISSN, publisher must send a request to the appropriate agency. ISSN has been standardized by ISO (ISO 3297:1998).

The International Standard Book Number (ISBN ISBN, 2004) is an intelligent and machine-readable identification number used to unambiguously identify publications. It was approved as ISO standard in 1970 (currently ISO 2108:1992); it does not provide central repository nor searching tools. ISBN is always associated to the whole publication. To obtain a ISBN publishers must pay a minimal fee. Many other digital resource identification standard exists, such as ISRC (SICI, 2004) and ISAN (ISAN, 2004). A detailed and complete description of the most used systems is provided in Vitiello, 2004.

#### **5. Normative documents identification: the NormeinRete approach**

Each law contains several references to other laws: one of the main problems dealing with normative documents arises from the need to retrieve all related documents. Users have to gather all the norms cited in order to correctly

interpret a law and to retrieve all subsequent laws that affected the original one in order to know the complete text that is “in force”.

The whole legislative corpus can be represented as a directed disconnected graph, in which laws are vertices and references among laws represent edges, defined through natural language expressions. The graph is directed because laws promulgation dates introduce a direction on the edges: a law cannot reference subsequent laws. The graph is disconnected because there are laws without references to others (for example the Constitution). Nowadays, most institutions provide systems to have access to norms by means of the Internet. Thus, nodes of the law graph are available as web resources and the hypertext of the connected part of the graph can be built using the usual web link mechanism based on resource physical address (URL). However, this approach has many drawbacks, that make it inefficient.

First of all, heavy manual activities are required, usually carried out through the following steps:

- recognizing all natural language expressions that are references to other norms;
- finding the physical address (i.e. URL) of the resource corresponding to the referred norm;
- marking-up the references inserting the physical address of the resource to be linked.

The need to modify the source document to insert the URL addresses in the appropriate locations reduces the overall document quality, because of the errors that can be introduced through manual typing. Moreover a norm could cite other norms not available on the web yet or the links inserted can become broken after a while. Thus a continuous monitoring is needed to ensure proper links behavior.

In order to build an effective and maintainable system supporting navigation through the graph a persistent, location-independent unambiguous identifier for nodes is required. This paper proposes a framework based on (i) an identifier named legalURN derived from the URN and based on (ii) a resolution system to resolve logical identifiers into physical addresses that is similar to the DNS used to resolve the self-explaining web sites names into numerical IP addresses. The mechanism used to build legalURN makes it an intelligent identifier. We illustrate legalURN syntax and semantics in Section 6 and Section 7.

The opportunity to automatically build an hypertext among laws relies on the following two considerations. The natural language expressions that refer to a norm usually contain repetitive patterns, making references automatically detectable through a specialized parser. Moreover, the URN is based on data

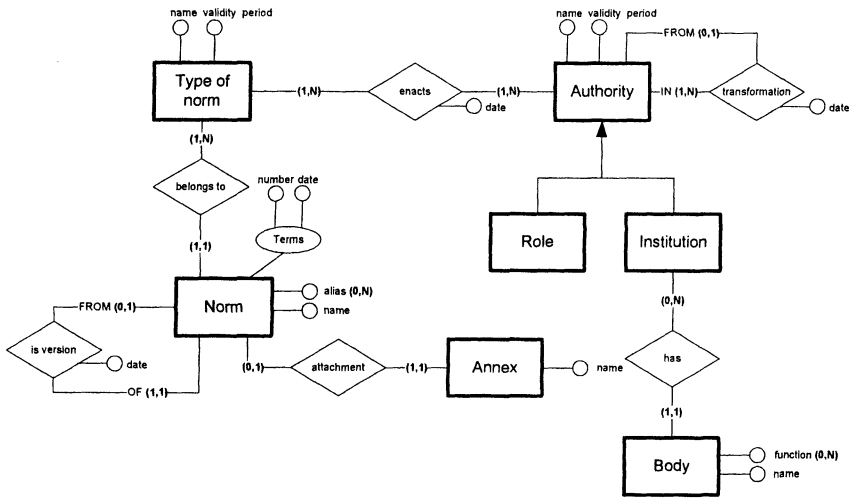


Figure 1. Entity-relationship model representing norms and related information

always included in the references and it can be built automatically. Thanks to these features, the legislative hypertext building process can be automatically performed associating the corresponding legalURN to each reference in the law. This function can be performed on the documents prior to publishing or on the fly. The parser that detects references in the text has a very high success rate (ranging from 95% to 97%).

To ensure the effectiveness of this approach some services are needed; these services provide the functionalities needed to create correct legalURNs and to resolve them. Thanks to this service, legalURN ensures actionability and persistence. We will describe the service infrastructure in Section 8.

## 6. LegalURN syntax

The legalURN consists of a combination of several elements. Figure 1 depicts the conceptual model on which we base our analysis, using the entity-relationship (ER) formalism. Each entity in the model has been mapped into a syntactical element. The most relevant elements are: the promulgating (enacting) authority, the type of the norm (law, decree, etc.) and the terms (i.e. date, number, etc.). The model allows to represent a set of more detailed specifications when needed, such as validity dates and alias.

The legalURN is compliant to the URN syntax. The set of rules to build a well formed legalURN has been defined by means of a formal grammar described by the Backus-Naur form. The grammar satisfies the following requirements: (i) each legalURN univocally identifies a norm and can distinguish

different versions due to subsequent modifications; (ii) the rules are easy, unambiguous and self-explanatory; (iii) legalURNs can be built automatically.

For the sake of simplicity, we illustrate the structure of legalURN focusing our analysis to the meaning of the elements. The complete grammar, allowed characters and other rules are reported in AIPA, 2001.

The general structure of a legalURN, as defined in the current version of the standard, is the following:

$$\langle \text{URN} \rangle ::= \text{urn} : \langle \text{NID} \rangle : \langle \text{NSS} \rangle$$

$\langle \text{NID} \rangle$  denotes the namespace identifier and assumes the arbitrary chosen value *nir*, that represents the acronym for *Norme in Rete*. Initially thought for the Italian context, the legalURN can handle different national legislative corpus introducing an element identifying the state where the norm has been enacted. A possible structure could be:

$$\langle \text{URN} \rangle ::= \text{urn} : \langle \text{NID} \rangle : \langle \text{STT} \rangle : \langle \text{NSS} \rangle$$

where element  $\langle \text{STT} \rangle$  should contain the ISO 3166 ISO, 2004 three letter code identifying the state. For instance, DEU for Germany or ITA for Italy.

$\langle \text{NSS} \rangle$  is the specific part of the namespace. It is defined as follows:

$$\langle \text{NSS} \rangle ::= \langle \text{document} \rangle [ @ \langle \text{version} \rangle ]$$

The element  $\langle \text{document} \rangle$  consists of three mandatory elements (promulgating authority, type of norm and terms) and of optional element (annexes). In most cases, these elements are sufficient to identify the norm.

$$\langle \text{document} \rangle ::= \langle \text{authority} \rangle : \langle \text{type} \rangle : \langle \text{terms} \rangle [ : \langle \text{annex} \rangle ]$$

The  $\langle \text{authority} \rangle$  element represents the public institution promulgating the norm and contains its name.

$$\begin{aligned} \langle \text{authority} \rangle & ::= \langle \text{authority-name} \rangle * [ + \langle \text{authority-name} \rangle ] \\ \langle \text{authority-name} \rangle & ::= ( \langle \text{institution} \rangle * [ ; \langle \text{body} \rangle ] [ ; \langle \text{function} \rangle ] ) | \langle \text{role} \rangle \end{aligned}$$

If several institutions enact a norm jointly, they must be all included; for example *ministry.finance+ministry.justice*. The element  $\langle \text{body} \rangle$  and the element  $\langle \text{function} \rangle$  can be used when the norm is promulgated by a department or a particular body (e.g. regional assembly, ministry department executive) of a institution; for example, *ministry.finance;department.taxes* or *ministry.finance;department.taxes;executive*.

The  $\langle \text{type} \rangle$  element represents the typology of a norm; for example: *law*, *constitutional law*, *decree*, etc.

$$\langle \text{type} \rangle ::= \langle \text{typology} \rangle * [ ; \langle \text{specification} \rangle ]$$

The optional element  $\langle \text{specification} \rangle$  contains fine granularity detail of the norm, as the common name of the norm or its classification. For instance, *law;remission.official.taxes*.



The `<terms>` contains the promulgation date and a unique number. Its syntax, very simple and intuitive, is the following:

```
<terms> ::= <dates>;<ids-norm>
<dates> ::= <date>*[, <date>]
<ids-norm> ::= <id-norm>*[,<id-norm>]
```

The number is usually that one associated with the act prior to its publication on the Official Journal. This number, in conjunction with all the other element's tokens, ensure uniqueness. However it must be taken into account that there are some kind of normative acts that have not an official number. Since there can be several acts of the same type issued by the same authority in the same date, an additional element is needed to ensure the unambiguousness of the legalURN. Therefore an internal unique number is associated to un-numbered acts when they are submitted for registration. This service is described in Section 8. These are some examples of the element `<terms>`: 2000-12-06;126, 1999-12-30,2000-01-13;p-2.

The last part of the `<document>` element is the `<annex>` element. It must be used when legalURN is built to identify a document annexed to a norm. Its syntax is the following:

```
<annex> ::= <id-annex>*[;<specification>]
```

Element `<version>` allows to manage different version of the same law due to subsequent changes. It contains the date of the last modifying norm that has provided modifications to be inserted in the original text. The element syntax is:

```
<version> ::= <norm-modifier-date>|inforce
<norm-modifier-date> ::= <date>
```

For example, `ministry.justice.law:1961-11-15;12@1987-06-03` identifies the law number 12 enacted by the Ministry of Justice in November 15, 1961, updated with the changes introduced until June 3, 1987. Since each element is meaningful, legalURN is a so-called intelligent identifier. Information contained in the legalURNs can be exploited in order to deeply analyze the links among norms, allowing to implement applications for legislative reorder support and laws classification.

## 7. LegalURN semantics

In the previous section we have illustrated the legalURN limiting our attention to element definitions and providing some information about what they contain. Nevertheless syntactic rules alone does not ensure the building of meaningful legalURNs, i.e legalURNs usable in our infrastructure. For example, the following legalURNs in the Italian namespace are not “semantically” valid:

Table 1. LegalURN: rules used to normalize expressions

Normalization rules
for each expression, must be considered only nouns and adjectives, discarding all connectives, articles, etc.;
type of authority must be in first position;
nouns must come before adjectives (as usually done in italian);
spaces must be replaced by dots;
acronyms must be expanded;
abbreviations must be expanded;
generic terms must be moved at the end;
all words must be in lowercase.

```
urn:nir:aaaaa:bbbb:1000-1-1;1;
urn:nir:ministry.defense:constitutional.law:1978-4-10;142;
urn:nir:ministry.finance:decree:2003-1-12;24;
```

These legalURNs are valid against the syntactic rules; nonetheless the first probably does not represent any law, because the elements have not any meaning. The second one has an internal inconsistency; in fact in Italy none but the Parliament can enact constitutional laws. Finally, the third one has a conflict between the authority name and the promulgation date, because Italian Finance ministry has been incorporated into the Treasure ministry in 1999.

Having regard to these aspects, two main issues can be pointed out:

- most of the legalURN elements can assume only a well defined set of value; for example, in the <authority> element only valid authorities names can be inserted.
- The value of each element can affect allowable values in other elements.

Internal consistency constraints, such as “norm typology must be consistent with norm date”, cannot be expressed directly in the Figure 1 ER model. Moreover, since syntactical elements represent real world entities, only a limited set of valid values can be used. Each entity value needs to be *normalized* before it can be used, in order to maintain the effectiveness of the resolution system and to limit inconsistencies. In order to avoid this problem, some registries containing the up-to-date information about allowable values and a service to normalize names are provided.

Concerning normalization, the standard proposes the rules of Table 1 in order to normalize expressions: For example, the Ministry of Treasure becomes “ministry.treasure” and EU is expanded in “union.european”. This approach is only one of the possible but other normalization mechanisms or rules may be used to represent names.

Some sets of values are more complicated than a simple list. In fact, it is necessary to take into account the “temporal” evolution of such values and their relationships. For example, while in the past in Italy there was a Ministry of Finance and a Ministry of Treasure, a few years ago the Ministry of Economy and Finance has taken in both. Thus, if one wants to register a legalURN for a norm of the Ministry of Finance with a year of promulgation equal to 2004, the system fires an alert message. Moreover, departments or offices often change denomination or are removed. Therefore, we need to record for each institution the name of its departments that have autonomous regulatory power, and the temporal validity of the name.

The legalURN system can discover inconsistencies and syntactical errors thanks to the presence of a centralized registry service. Registries can be used both to register legalURNs and to correctly insert persistent references to norms.

## **8. The NormeinRete system architecture**

In this section we describe the legalURN management system implemented within the context of the NormeinRete project. Although at this stage the system adopts a simplified architecture, the successful implementation shows the benefits deriving from the proposed approach.

The system architecture aims to cover all the aspects concerning legalURN life-cycle, providing functions to easily manage and maintain the standard. The overall architecture is depicted in Figure 2. The main architectural elements are the registration services, the central registries and the legalURN resolution service. Currently, these modules are implemented in a centralized manner and are available through the NormeinRete site. Nonetheless, the modular architecture easily allows to implement an equivalent distributed system.

Administrations have two ways to share their normative data:

- 1 register each norm using the registration service;
- 2 become a NIR node, leaving to the resolution module the task to automatically indexing the norms.

In the second case, the administration legal documents must be formatted and saved with specific file system settings.

### **8.1 Registration services**

The registration service puts at the users’ disposal two functions: (i) a function to register a new legalURN or associate a new resource to an existing legalURN, and (ii) a function to register an institution as NIR node. In the former case, users actively interact with the system; whereas, in the latter case

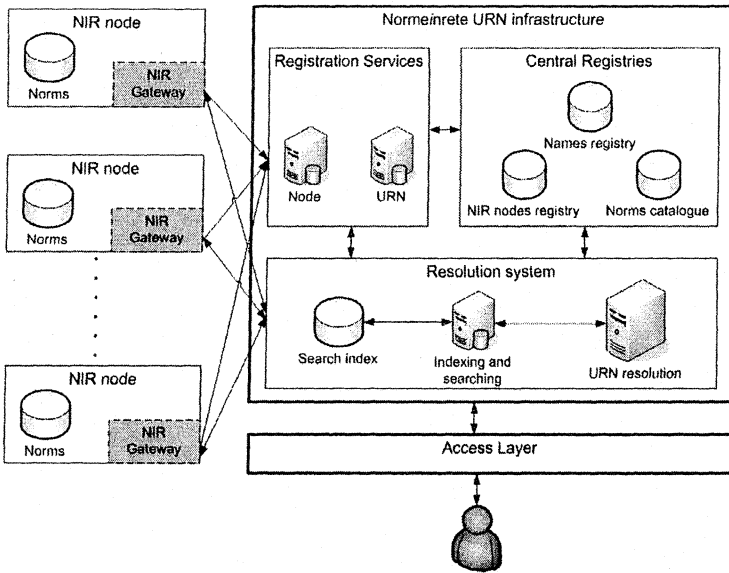


Figure 2. The NIR legalURN system architecture

the system is responsible to load a new document on the research index and to build new legalURNs.

We think that the responsibility for legalURN registration should belong to the institution enacting the norm. In fact, it precisely knows the information needed to build the legalURN, thus avoiding mistakes or incomplete data. This rule requires a well defined inter-institutional organizational asset.

When an institution wants to make a document available by means of legalURN mechanism, it must register it on the system. At this stage additional metadata must be inserted. The system turns the inserted information into a legalURN consistent with the rules illustrated in Sections 6 and 7 and joins it to the physical resource. If the obtained URN is already present, the system associates the new physical resource to it. Beginning from this moment, the URN can be used to link a norm, using the resolution module. The institution that registers a norm must keep its physical resource aligned with the ones they have associated with the URN, in order to avoid errors in the resolution phase.

## 8.2 Central Registries

The central registries store information needed to allow effective standard management. The registry of official authority names is needed in order to guarantee consistency in the construction of legalURN. The registry contains the names of the institutions, name history with the periods of validity, relationships with other institutions (such as derived by, substituted with, joined

in, etc.). The NIR-nodes registry contains all the information needed to allow automatic interaction between NIR agents, such as spider agent and data retrieving agent, and the application gateways located in each domain. Finally, the norms catalogue contains for each legal document a minimum set of metadata such as the title, a basic classification, associated legalURN and the list of known physical resources (such as internet addresses) where the versions of the corresponding document are published. Metadata can be used to implement more advanced services.

### **8.3 The legalURN Resolution service**

The resolution mechanism is quite similar to the Internet domain name system (DNS). Thanks to this module, legalURN becomes an actionable identifier. Resolution service takes as input a legalURN and returns, if it exists, the related resource(s), e.g. web page or file associated to it. Currently, the system architecture is fully centralized, with one node that manages all the requests. However, the resolution system can be built both as centralized and decentralized with a hierarchical structure. We think that the architecture should be distributed, and each state or local authority should maintain its own registries and resolution services. Like DNS, the system may exploit the legalURN structure to simplify the operation and to quickly direct the search toward the right national or local level node of the search tree. Reading the <NID>, system can easily verify if the norm belongs to its domain or the request must be forwarded to another node.

Despite DNS, where the relation between the logical address and the IP address is many to one, in our case more resources can exist for the same legalURN; in other words the relation between the legalURN and the physical resource is one to many. This complicates the resolution mechanism, because the system needs a decision phase. In order to solve this problem, we are investigating a double level resolution mechanism, which allows users to specify the preferred version using a set of simply indicators, such as data source completeness or trustworthiness. This approach is similar to the one proposed in Mecella et al., 2002.

## **9. Conclusions and future works**

We have described the identification system for legal digital resources implemented within the Italian project NormeinRete. The proposed mechanism is based on an identifier compliant with the URN standard, named legalURN. The identifier has some important properties: it is actionable, intelligent and persistent. The system is currently working limited to the Italian legislative corpus, but can be easily extended to cover general legal documents. The implemented system has shown the effectiveness of our approach. Moreover,

other European institutions have shown interest in legalURN. It is important to underline that the legalURN standard belongs to a complete legal document framework, which is able to manage the complete life-cycle of a legal document.

We are working to refine the system architecture of legalURN in order to simplify extensibility and scalability to support the standard adoption in a wide range of contexts. Many open research lines are currently being investigated; among all, a more formalized definition of the semantics rules, using logic formalism, and the introduction in the identifier of information about the document classification with respect to a semantics schema or a ontology.

Representing a legislative corpus as a graph will allow to exploit some well known graph properties and algorithms to obtain important practical results. We are working to develop systems that are able to support normative reorder, laws impact analysis and tools to simplify well done laws creation. For example, by analyzing the graph we can identify the parts of the graph characterized by an high degree of cohesion and collapse the nodes strictly connected into a unique node, thus supporting legislative reorder task. A more detailed description will be provided in a following paper.

## Notes

1. <http://www.ittig.cnr.it/Index.htm>
2. <http://www.cirfid.unibo.it/>

## References

- AIPA (2001). Definizione delle regole per l'assegnazione dei nomi uniformi ai documenti giuridici. Circolare n. AIPA/CR/35. In Italian.
- Boer, A., Hoekstra, R., Winkels, R., van Engers, T., and Willaert, F. (2002a). Proposal for a dutch legal xml standard. In Traummüller, R. and Lenk, K., editors, *EGOV 2002 Conference*, pages 142–149, Berlin. Springer.
- Boer, A., Hoekstra, R., and Winkels, R. (2002b). Metalex: Legislation in xml. In Bench-Capon, T., Daskalopulu, A., and Winkels, R., editors, *Legal Knowledge and Information Systems. Jurix 2002: The Fifteenth Annual Conference*, pages 1–10, Amsterdam. IOS Press.
- DOI (2004). The digital object identifier system. <http://www.doi.org/>.
- Grandi, F., Mandreoli, F., Tiberio, P., and Bergonzini, M. (2003). A temporal data model and management system for normative texts in xml format. In *Proceedings of the fifth ACM international workshop on Web information and data management*, pages 29–36. ACM Press.
- ISAN (2004). The international standard audiovisual number. <http://www.isan.org/>.
- ISBN (2004). The international standard book number. <http://www.isbn.org/standards/home/index.asp>.
- ISO (2004). Standard for country names. Available on line (link checked November 2004): <http://www.iso.org/iso/en/prods-services/iso3166ma/index.html>.
- ISSN (2004). The international standard serial number. <http://www.issn.org/>.

- Lupo, C. and Batini, C. (2003). A federative approach to laws access by citizens: The “normein-rete” system. In Traummüller, R., editor, *Proceedings of Second International Conference Electronic Government (EGOV)*, pages 413–416. Springer-Verlag.
- Masinter, L. and Sollins, K. (1994). Functional requirements for uniform resource names, rfc 1737. Available on line (link checked September 2004): <http://www.faqs.org/rfcs/rfc1737.html>.
- Mecella, M., Scannapieco, M., Virgillito, A., Baldoni, R., Catarci, T., and Batini, C. (Irvine, CA, 2002). Managing Data Quality in Cooperative Information Systems. In *Proceedings of the Tenth International Conference on Cooperative Information Systems*.
- Moats, R. (1997). Urn syntax, rfc 2141. Available on line (link checked September 2004): <http://www.faqs.org/rfcs/rfc2141.html>.
- Palmirani, M. and Brighi, R. (2002). Norma-system: A legal document system for managing consolidated acts. In Cicchetti, R., Hameurlain, A., and Traummüller, R., editors, *Proceedings of the 13th International Conference DEXA*, pages 310–320. Springer-Verlag.
- Paskin, N. (2003). *Digital Rights Management - Technological, Economic, Legal and Political Aspects*, chapter Components for DRM Systems: Identification and Metadata (2.3.1). Springer, Available on line (link checked September 2004): [http://www.doi.org/topics/drm\\\_paskin\\\_20030113\\\_b1.pdf](http://www.doi.org/topics/drm\_paskin\_20030113\_b1.pdf).
- SICI (2004). The international standard recording code. <http://www.ifpi.org/isrc/>.
- Sollins, K. (1998). Architectural principles of uniform resource name resolution, rfc 2276. Available on line (link checked September 2004): <http://www.faqs.org/rfcs/rfc2276.html>.
- Spinoso, P. (2001). Identification of Legal Documents Through Urns (Uniform Resource Names). In Signore, O. and Hopgood, B., editors, *Proceedings of the Euroweb 2001 Conference “The Web in Public Administration”*.
- van der Werf-Davelaar, T. (1999). Identification, location and versioning of web-resources. URI Discussion paper, available on line (link checked September 2004): <http://www.kb.nl/coop/donor/rapporten/URI.html>.
- Vitiello, G. (2004). Identifiers and identification systems. *D-Lib Magazine*. Available on line (link checked November 2004): <http://www.dlib.org/dlib/january04/vitiello/01vitiello.html>.