

The NIR Project

Standards and tools for legislative drafting and legal document Web publication

C. Biagioli, E. Francesconi, P. Spinosa, M. Taddei

ITTIG – Istituto di Teoria e Tecniche per l'Informazione Giuridica C.N.R. Firenze, Italia
e-mail: {biagioli, francesconi, spinosa, m.taddei}@ittig.cnr.it

Abstract. The national Project *Norme in rete* [Legislation on the Net] aims at making easier the retrieval and the navigation between legal documents in a distributed environment and to encourage the development of systems with characteristics of interoperability and effective of use. In order to obtain this, two standards have been defined: a URN standard, to identify these materials through uniform names, and XML-DTDs to describe legislative documents. In this paper the definition of such standards and the developments of tools aimed at making easier their adoption are illustrated. Particularly a cross-reference parser recognising texts and constructing the corresponding URN, and an editing system helping draft law texts and to handle legacy legislative documents according to the NIR-DTDs are described.

1. Introduction

Users increasingly feel the need to retrieve legal documents from the Web and the links between them in order to learn about the law and fully understand legal texts. In Italy a project, the *Norme in rete* (NIR), has been proposed in order to create a unique access point on the Web with search and retrieval services of legal documents. The project proposes to realize a system which, based on presently available technology and standards, is able to multiply the benefits each public authority can pursue in terms of information transparency and accessibility through the Web publication of documentation of prescriptive interest.

This paper is organized as follows: in Section 2 a brief description of the aims of *Norme in rete* project is reported; in Section 3 the standards established by the NIR project are described; in Section 4 the tools developed within the project to make easier the adoption of such standards are illustrated.

2. The *Norme in rete* Project

Since mid 1999, the AIPA [Italian Authority for Information Technology in the Public Administration], within the domain of development initiatives for the Unified Network for the Public Administration, finances and co-ordinates, in conjunction of the Italian Ministry

of Justice, an inter-sector project called *Norme in rete* (NIR) ["Legislation on the Net"], which aims at improving and simplifying the retrieval of legal documentation made accessible by different institutional bodies via the Internet.

The inter-sector projects are initiatives involving and encouraging the collaboration between various administrations and public bodies, also through the exchange of experience and knowledge, attempting to eliminate information historical fragmentation and, therefore, improving the provision of adequate services for citizens.

Basically *Norme in rete* aims at setting up a specialised portal, equipped by a search engine, for retrieving legal documentation, and at guiding users towards relevant sites of public authorities participating in the project. Moreover the portal provides a series of special services, such as, for example, news about particular matters, best practices, tools to support the public authority activities within the project.

Work on the project is directed along two lines: the first is aimed at analysing and exploring the technical aspects and at defining the common standards to be used; the second is oriented towards building operational prototypes in real environments, testing feasibility and designing the project basic architecture, with a view handing the development and management of the system over to a service company. They are aimed at constructing a federative service for public authorities based on common standards, without interfering with their computer systems. Hereinafter the results of these two lines of work are described.

3. The NIR standards

The feasibility study of the NIR project proposed the adoption of XML as a standard for representing legal documents. This study aimed at representing a legal text with respect to its formal structure, using also additional meta-information and a uniform cross-referencing system providing documents with characteristics of interoperability and effective of use.

This preliminar study, carried on by two specific national work groups produced two main official standards:

1. a standard for cross-referencing legal documents has been defined in accordance with the uniform name (URN) technique: an unambiguous identifier, that allows the references to be expressed in a stabile way, independently of their physical location;
2. a standard for legal document description has been formulated by defining XML-DTDs (NIR-DTDs) of increasing degree of depth in text hierarchy description for different kind of legal documents (similar initiative is the MetaLex project [1]).
As well as including the NIR-URN standard for cross-references, the NIR-DTDs provides:
 - a structural description of text, establishing constraints in the hierarchy of the formal elements of a legislative text (collections of articles);
 - a specification of the metadata which can be applied to a legislative document or to parts of it.

3.1 The URN standard

Within the NIR project, documents are identified through a uniform name. Uniform Resource Names (URNs) were conceived by the Internet community for providing unambiguous and lasting identifiers, independent of physical location, of network resources.

In legal documents, references to other legislative measures are very frequent and extremely important. The hypertext links of the Web meet this need, but do not appear to be suitable for wide-scale use in the law: reference to the resource referred to is, in fact, based on its physical location expressed in a uniform mode through its URL (Uniform Resource Locator), which presents the following well-known problems:

- difficulty in knowing the location of the cited resource;
- the loss of validity over time of the locations (URL) in the references;
- the impossibility of referring to resources that have not been published yet;

which, therefore, make the network of links between documents extremely limited with respect to their potential and to their increasing unreliability over time.

In order to avoid these problems, a system of references based on assigning a uniform name to each legal resource and on resolution methods (RDS: Resolver Discovery Service) able to retrieve the corresponding object has been chosen. These tools are in conformity with those defined within IETF (Internet Engineering Task Force) by the special working group (URN Working Group) and described in various documents – from the official standards (RFC: Request For Comments) to the drafts – to which alignment is guaranteed even in the future.

Assigning a uniform name to every legal document has the scope of associating every legal document with an unambiguous identifier, in standardised format, that only depends on the characteristics of the document itself and is, therefore, independent of on-line availability, of physical location and of access mode. This identifier is used as a tool for representing the references – and more generally every type of relation – between the legal acts. In an on-line environment with distributed resources between different Web publishers, its use facilitates the construction of a global hypertext between legal documents and a knowledge base storing the relations interconnecting them.

The association of the uniform name to the document occurs through metainformation, that may be:

- inserted in the document itself: it is the solution that can be adopted in HTML files (through the META tag) and also in XML files (through a suitable tag);
- external but strictly related to the document: by traditional techniques as a specific attribute in a database, or using growing methods as adopting RDF technology.

In any case, the software tools used must be able to implement and update the (distributed or centralised) catalogues which are functional for resolution and, therefore, to give access to the document through the uniform name. Other metainformation (for example, details, title, subject-matter, relations, whether in force, etc.) which enrich the system response, can be present in these catalogues that store the uniform name and location for each document.

The uniform names system of the domain of interest must include:

- a schema for assigning names capable of representing unambiguously any legal measure, issued by any authority at any time (past, present and future);

- a resolution mechanism – in a distributed way – from uniform name to on-line location of the corresponding resources.

The application of uniform names in the law, already proposed in ITTIG's *Feasibility Study*, was then defined in greater detail by the special NIR Working Group: the final result was contained in a document [2], which defines the perfected standard, a standard about to be adopted as a technical regulation by AIPA and transmitted to the Public Administrations through a specific circular (n.35, November 2001).

In conformity with RFC 2141 *URN Syntax* [3], which defines the general syntax of a uniform name, for legal documents a name-space identified by "nir" (this space identifies the context in which the names are valid and significant) has been defined and, therefore, the relative URN have the following format:

`<URN> ::= "urn:nir:" <NSS-nir>`

The specific name `<NSS-nir>` must contain information appropriate for univocally identifying the document. In the legal domain they are essentially four data: the enacting authority (or the authority referred to), the type of measure, the details and any annex. For legislation, it is also necessary to distinguish between any later versions of the document, following amendments that have been made over a period of time. In this case, the identifiers of the legislative act remain the same, but information is added regarding the version under consideration. Therefore, the more general structure of the specific name appears as follows:

`<NSS-nir> ::= <document> ["@" <version>]`

A structure for identifying the document is defined, composed of the four fundamental elements mentioned above, clearly distinguished one from another in accordance with an order identifying increasingly narrow domains and competence:

`<document> ::= <authority> ":" <measure> ":" <details> [":" <annex>]`

The main elements of the uniform name are generally divided into several elementary components, each having established rules of representation (criteria, modes, syntax and order). Such a syntax allows the automatic construction of the URN, starting from the text of the citation. The complete syntax specification of the uniform names belonging to the "nir" name-space can be seen in [2] and in [4], whilst some important examples of uniform names of legal documents are:

Act 24 November 1999, No. 468

`urn:nir:stato:legge:1999-11-24;468`

Decree of Ministry of Finance of 20.12.99

`urn:nir:ministero.finanze:decreto:1999-12-20;nir-3`

AIPA circular of 21 June 2001, No. 31

`urn:nir:autorita.informatica.pubblica.amministrazione:circolare:2001-06-21;31`

Decision of the Italian Constitutional Court No.7 of 23 January 1995

`urn:nir:corte.costituzionale:sentenza:1995-01-23;7`

To each uniform name, the system of resolution has the task of associating the respective network locations. It is based, within a distributed architecture, on two basic components: a chain of information in DNS (Domain Name System) and a series of resolution services from URNs to URLs, each competent within a specific domain of the name space. Particular attention has been paid to the resolution system in order to provide an answer to the user, even in case of uncompleted or uncorrected uniform names, derived from uncorrected

The NIR Project

Standards and tools for legislative drafting and legal document Web publication

citations (for example the resolution service gives back the list of the documents whose URNs partially match the provided URN, or it attempts to correct automatically the URN itself).

3.2 The NIR-DTDs standard

As well as the NIR-URN standard, the NIR project has defined a standard based on XML, aimed at describing the content of legislative documents. For this purpose three DTDs with increasing degree of depth have been established:

- the “DTD flessibile” (*nirloose.dtd*) contains about 180 elements: it does not establish any mandatory rules (unless in a very small quantity) and it is used for legacy legislative documents not following drafting rules;
- the “DTD base” (*nirlight.dtd*) contains about 100 elements: it represents a subset of the “DTD completo”: it is useful to train users in adopting the DTD standards;
- the “DTD completo” (*nirstrict.dtd*) contains about 180 elements: it follows legislative drafting rules and it is used to write new legal documents.

The “DTD flessibile” and “DTD completo” are composed by four common files:

1. *global.dtd*: containing general definitions;
2. *norme.dtd*: containing definitions of the division structures;
3. *text.dtd*: for text, table and form structure definitions;
4. *meta.dtd*: containing metadata schemes definitions. Differences are present in the

main files *nirstrict.dtd* and *nirloose.dtd*.

The *nirstrict.dtd* establishes an order to the partitions of a law text. Collections of articles are still considered the basic elements of the norm (their numbering is independent from the hierarchical organization of the other elements). Numbering of the divisions is mandatory. Titles of the divisions are not provided, while they are optional for the other elements.

The *nirloose.dtd* establishes only few constraints and it is used for legacy legislative documents which usually do not follow particular legislative drafting rules.

The NIR-DTDs basically describe a legislative text under two profiles:

- the *formal profile* which considers a legislative text as made up of *divisions*;
- the *functional profile* which considers a legislative text as composed by elementary components called *provisions* (fragment of a regulation) [5][6][7].

In other words, the fragments of text inserted have a formal and a functional appearance. They are, at the same time, partitions and provisions, according to whether they are seen from a formal or fundamental view-point. The two points of view can be alternated as required during the definition of the text.

In particular the functional profile can also be considered as composed by two sub-profiles: the *regulative profile* and the *thematic profile*. The first one reflects the lawmaker directions, the second one the peculiarities of the regulated field. On the NIR-DTDs point of view, the regulative profile is identified by particular metadata called *analytical provisions*, the thematic profile are partly illustrated in the so-called *subjects* of the provisions.

4. The tools to support NIR standards

In order to make easier and faster the adoption of such standards, some tools have been developed within the NIR project:

- 1 A cross-reference parser, aimed at recognising, within a law text, the cross-references to other law texts and constructing the corresponding URN [4].
- 2 An editing system to help draft law texts, to qualify documents with metadata, to handle legacy legislative documents according to the NIR-DTDs. The editing system is designed to implement also the main lawmaking rules and the usual procedures of law drafting [8][9].

4.1 The cross-reference parser

The uniform name for cross-references will be used on a large scale for legislative documents with regard to references as a value of the HREF attribute of the hypertext link with the document referred to. This link can be created in two ways:

- by manually inserting, in the referring document, the link with the uniform name: this is possible simply on the basis of the citation text itself, however it is a burdensome operation especially for documents already on-line;
- by electronically constructing (either permanently or temporarily) the link with the uniform name, through parsers of references within the text: a more economic operation but subject to a certain percentage of errors, due to the fact that references are not always precise or complete. This solution could, nevertheless, be acceptable for already published documents.

In any case, whatever the means adopted, newly produced documents in XML format in the NIR environment must, in conformity with the relative DTD, express references through the uniform name of the document referred to.

In our project a cross reference parser has been developed as support to URN implementation and it is available both within the editing system and as Web service.

Developed by ITTIG (reachable from www.normeinrete.it), it is able to construct the uniform name associated with the legal measure found in a reference. It can also be used by submitting a text (containing one or more references) to analysis:

- typed-in by the user or captured directly from a document displayed on the screen, with the copy/paste function (eventually even the entire document);
- by uploading a local file;
- indicated through its location on the Web (URL).

The parser generates a hypertext link (currently containing the uniform name of the main classes of cited legal measure) in connection with each reference found in the text submitted to analysis, with the possibility, therefore, of retrieving the resources associated to this (by using the above mentioned resolver). This module is based on a grammar implementing a bottom-up parsing strategy. The service is able to parse documents of different format (TXT, HTML, XML), and to give back the list of the document URN, or a parsed document where cross-references are marked as HTML links or by the related

The NIR Project

Standards and tools for legislative drafting and legal document Web publication

XML-NIR mark-up. Moreover, on user demand, this service is available within the window where the documents, given back by the NIR search engine, are displayed.

4.2 The NIR Editor

In order to help the drafting of law texts according to NIR-DTDs standard, a specialized editor for NIR documents has been developed.

Even though programs for drafting texts in XML already exist, we have decided to develop a specific environment to handle NIR-XML documents. The limits of present XML editors in fact lie in their newness and the inadequate development of their editing functions compared with current word processing programs. Even though this limitation will soon be overcome by their normal evolution, however another limit persists with respect to reasonable user expectations: it concerns the generality of such editors which have been created to handle any type of document; such editors, therefore, lack of particular functions, which help the user to compose a legislative text. In our opinion this justifies the initiative of realizing a writing environment for NIR-XML documents.

Moreover, whilst generic XML editors are limited to detecting and preventing descriptions which are inconsistent, the editor specializing in prescriptive texts can offer proposals by choosing to insert appropriate descriptions. This is the main difference between a generic editor and an ad hoc application such as the one we are proposing. The passive act of simply verifying the correctness of what has been drafted is replaced by constructive action on the part of the editor which intervenes in the creation of the text by suggesting, helping and often taking the place of the draftsman.

Specific NIR functions, therefore, have been developed to handle legislative documents; in conjunction to a law drafting environment they compose the NIR Editor.

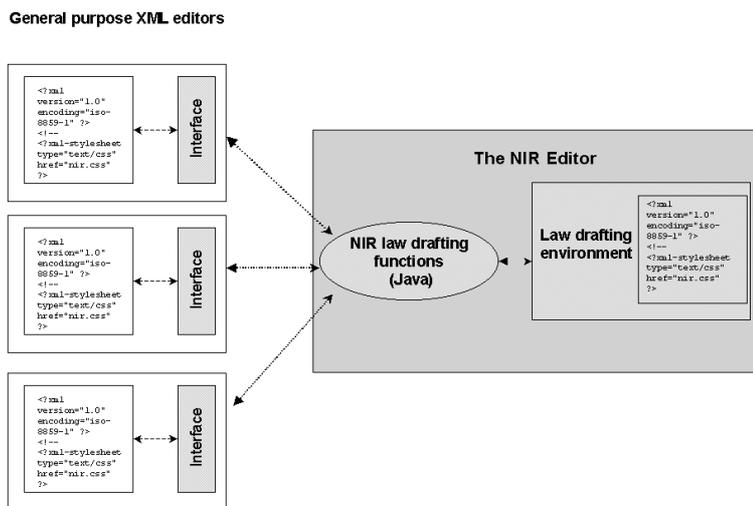


Fig. 1 The NIR Editor and its connections to general-purpose XML editors

The architecture of the NIR Editor is represented by a kernel of Java specific NIR functions library, fully integrated within the NIR law drafting environment; they can also be integrated to the main XML general purpose editors supporting a Java API (Fig. 1).

The NIR Editor operates within the URN and DTD NIR framework and is designed to assist the drafting of new texts according to that standard, as well as to process existing texts in any format. Two working situations are thus catered for: the processing of an existing text or the processing of new texts, with its different situations: composition and organization of new texts.

4.2.1 Importing texts

In this case, instruments for recognizing the basic aspects of the texts are available, which allow automatic pre-marking of all the parts of the structure recognized in the text analyzed, in accordance with the DTD relevant to the measure being examined, thus recognizing the formal profile of the legislative text.

This *structure parser* includes also the *cross-reference parser* and it is designed to help the XML conversion of documents which otherwise would have to be carried out completely manually. However, the recognition functions are also useful for the conversion and marking of simple passages which already exist or have been drawn up in non XML environments, to be inserted in texts in the process of being drafted.

Currently the structure parser implements a non-deterministic finite-state automata (NFA), where the states are represented by the elements of the NIR-DTD *nirstrict.dtd*, and the transitions among the states are associated to formal rules of document parts division.

Normal editing instruments are available for improving the document as well as for carrying out revisions and making notes (meta-data). The completion and correction of the pre-marking itself is possible in the drafting environment, using all the functions available for the environment in question, from the validity of the document to sequence controls on the numbered divisions, functions which make preparation even faster and safer.

The result of the structure parsing function is the formal profile of the text which is established by the structural elements of the NIR-DTDs.

A further way of marking a pre-existing text is represented by the application of the metadata to a law text, therefore the recognition of the functional profile of a legislative text, whose NIR specification within the NIR-DTDs is established by *meta.dtd*.

As the marking of the formal structure, it can be manually carried out, however this function can be particularly time consuming. Therefore, within the NIR Editor a module supporting the user in *provision classification*, based on machine learning techniques for text classification [10], is currently under implementation. It extracts automatically from the text of the provisions their relevant meanings according to the NIR analytical metadata schemes. Particularly a naïve Bayes approach of text classification is being tested.

Similarly a module of *extraction of the provision subjects* based on natural language processing techniques is under implementation.

4.2.2 The composition of new texts

For the *composition* of new texts, dialogue windows, with internal variants according to the type of deed, facilitate the insertion of the data relative to the same located in the initial and

final part and, if considered necessary, also permit the introduction of the metadata provided for by the NIR-DTDs in this part of the document.

It is then possible to draft the body of the new act with the help of various functions which permit easy text structure construction and management. The automatic numbering function of the divisions represents the case of greatest delegation of initiative to the editor. The draftsman can completely ignore the problem of the correct sequences of divisions inside the ranks of the collection of articles. The editor is concerned with the question as a whole. Even in the event of text movements or variations, the re-numbering function is able to keep all the sequences coherent, as well as the linked internal quotes.

Automatisms are present as far as the construction of external and internal cross-references are concerned, as well as the transformation of the same into links. There is a guided composition window for the formulation of cross-references as well as instruments of automatic recognition and construction of references via URNs; moreover, the measure referred to can be checked on the network, integrating the URN resolution service into the drafting environment. As with word processing programs, it is possible to start by determining the structure of the text and insert the content of the various parts afterwards, or else passages can be inserted in no particular order, then organized and inserted into a suitable structure at a later time.

If the draftsman is clear about the structure of the text from the beginning, he will find it more suitable to formulate the whole thing using the special functions and subsequently insert the content of the text divisions (paragraphs, letters and numbers). Vice versa, in other cases, the text could be the result of a disorderly accumulation of fragments which will be subsequently organized.

During the composition, a further valorization of a legislative text is represented by the application of the analytical metadata to the divisions. Once a division has been drafted, the user can, if he wishes, locate the terms contained in the same corresponding to the so-called *subjects*, qualify them by comparing them with those provided for by the NIR-DTDs (provisions-subjects menu), and mark them. These notes are understood by the NIR-DTDs to be metadata and associated with the division in the electronic legislative document.

Therefore, in the event that metadata have been inserted which are the result of documentary requirements, it is possible to make use of these notes to determine the logical structure of the text being processed, as well as for subsequent network information searches. The editor will use the same criteria at the basis of the creation of the new associations to suggest possible titles (or headings) for the same. In practice, this consists of the clarification of the criteria of association followed in the form of significant terms for the titles of the various associations constructed.

In the future, a morphological-syntactical module is planned for the generation of titles in natural language, complete with articles, prepositions etc.

4.2.3 *The organization of new texts*

For the *organization* of new texts, two alternative strategies can be followed:

- the *formal strategy*;
- the *functional strategy*.

The *formal strategy* considers the text according to the formal profile: the text is made up of *divisions* (collection of articles). Using the formal strategy the partitions to be organized are chosen by the draftsman. He will indicate the provisions of similar rank to be

grouped together as he goes along and the editor will create new associated parts of an immediately higher rank, applying the rules of formal text structuring to the same. If, for example, the divisions, which are combined together into a new division are paragraphs, the new division container will be an article.

The *functional strategy* considers the text according to the functional profile: the elementary component of a text is a *provision* (fragment of a regulation). The draftsman carries out the same operations in an indirect way: the partitions to be organized are chosen according to their content, affinities etc. as well as it is decided where they should be placed in the text, observing the customary procedure of presentation used in some (few) rules of legislative technique.

In other words, the fragments of text inserted, obviously with proper meaning, have a formal and a functional appearance. They are, at the same time, paragraphs and provisions, according to whether they are seen from a formal or fundamental view-point. The two points of view, both completely compatible with each other, can be alternated as required during the definition of the text.

The attention to the functional profile of a legislative text based on metadata is one of the key points of the NIR Editor; this is the precondition of creating at least a domain-specific semantic portion of the Web.

The functions implemented in the NIR editor prototype are available at the address:
<http://www.idg.fi.cnr.it/organizzazione/personale/biagioli/Legal-drafting.htm>

5. References

- [1] A. Boer, R. Hoekstra, R. Winkels, "MetaLex:Legislation in XML", in Proceedings of JURIX 2002: Legal Knowledge and Information Systems
- [2] P. Spinosa. "Assegnazione dei nomi uniformi ai documenti giuridici", Novembre 2001.
- [3] R. Moats, K. R. Sollins, *URN Syntax*, RFC 2141, May 1997.
- [4] P. Spinosa, "Identification of Legal Documents through URNs (Uniform Resource Names)", in Proceedings of the EuroWeb 2001, *The Web in Public Administration*
- [5] Biagioli, C. (1997b), Towards a legal rules functional micro-ontology, *Proceedings of workshop LEGONT'97*, Melbourne, 1997.
- [6] Valente, A. e Breuker, J. "A Functional Ontology of Law", C. Ciampi, F. Socci Natali, G. Taddei Elmi (eds), *Verso un sistema esperto giuridico integrale*, CEDAM.
- [7] van Kralingen, R.W. (1995), *Frame-based Conceptual Models of Statute Law*, Kluwer Law International.
- [8] C. Biagioli, "An XML editor for Legislative drafting", JURIX Workshop on e-Government, 2002
- [9] Branting, L. K., J.C. Lester, C.B. Callaway, (1997), "Automated Drafting of Self-Expalinig Documents", *The Sixth International Conference on Artificial Intelligence and Law*, Melbourne, Australia.
- [10] F. Sebastiani. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1): 1-47.