

Linking Legal Sources in a Shared Web Environment

E. Francesconi, G. Peruginelli
ITTIG – Institute of Legal Information Theory and Techniques
Italian National Research Council,
Florence, Italy
Tel: +39 055 4399600
Fax: +39 055 4399605
{francesconi, peruginelli}@ittig.cnr.it

Abstract

The functionality of the Portal to legal literature, which is under development at ITTIG, is described, focusing on a linking mechanism created to allow navigation between statutory rules, judicial cases and legal literature pieces of work. Reference is made to previous works carried out by Italian institutions for the identification of legislation and case law resources through URN identifiers. The design of a URN for legal literature material is planned to ensure cross-access to legal sources on the basis of user surveys carried out for the purpose of better customize the Portal's services. The main features of the proposed URN for legal literature material are described as well as tools developed for the Portal based on an OpenURL resolver mechanism in order for users to start from the examination of a law (legislation) to its interpretation (legal literature) and further to analyse the enforcement of that law in judicial cases (case law).

Keywords: metadata for legal resource identification; URN; automatic cross linking; OpenURL; OAI-PMH.

1. Introduction

Accessing legal information is a primary requirement for a variety of communities: ordinary citizens, scholars and professionals. The dissemination of legal information contributes to the rule of law and to the overall ideals of democracy in a number of ways. The legal world is on the brink of a fundamental transformation that is being precipitated by existing and emerging information technologies. Legal practice and administration of justice will be profoundly changed in ways that all citizens will greatly benefit from such evolution. Information technology has opened up new methods of publishing law and official information; in particular the wide and ever increasing Internet availability has reinforced demands that law should be made freely available on this medium (1).

In particular legal literature, although complementary to statutory rules and judicial cases, is an essential part of legal research, and its specific function is to contribute to the interpretation and application of law, thus supporting the activity of a wide category of users, like scholars, judges and lawyers, administrators, politicians, reference librarians and information professionals, as well as ordinary citizens.

Under the pressure of massive and rapid production and delivery of information content, users demand improved, harmonized and fast access to a wide range of legal literature resources, asking at the same time for quality control.

In this direction the Institute of Legal Information Theory and Techniques of the National Research Council (ITTIG-CNR) has planned a Portal to legal literature, putting efforts in

developing a system to ensure a unified point of access to legal literature (2) (3). The architecture of the Portal is designed to integrate in a unique view both data coming from structured repositories and unstructured Web documents, using Dublin Core metadata set. So far the project's focus has been mainly on implementing and enhancing semantic search facilities to retrieve legal material through metadata.

At the moment, to ensure effective navigation facilities to users, developments are underway, to provide linking techniques and retrieval tools allowing users to navigate across different types of legal sources (legislation, case law and legal literature). Retrieval and browsing are two main tasks to be provided in an information retrieval system. In fact, users are accustomed to combine these two tasks to refine or expand their search. In particular the *browsing task* becomes important when resources are natively interrelated to each other through references, as it occurs in the legal domain (Section 3). "Native" interrelation among documents means references explicitly expressed within the texts like textual citations or bibliographic references. Through these links a text of law is analysed, its interpretation evidenced as well as its enforcement.

This has confirmed the importance for legal users of accessing statutory rules, judicial cases and legal literature in a way that these sources be interrelated so to allow users navigate across them.

This paper is organized as follows: in Section 2 the scenario of legal information access in Italy is discussed; in Section 3 the linking scenario of different legal sources and the need of intelligent identifiers is underlined; in Section 4 and 5 identifiers for Italian legislation and case law are shown; in Section 6 a proposal for legal literature identification is illustrated. Finally, in Section 7, some conclusions are drawn.

2. A glance to legal information access in Italy

Italian legal information is mainly produced and delivered on print formats, but the increasing amount of legal texts output has brought about a crisis in information access and the traditional tools are no more sufficient to its distribution. Efforts are underway to use new technologies to overcome this problem, increasing the provision of electronic resource on the Web. In this context the integration between traditional and electronic sources, both off line and on line, is becoming a major requirement. In Italy the production of legal electronic resources started in 1995; initially this material was found only on academic and research institutes' Web sites. Only in 1998 legal information started to be offered on the institutional Web sites of the major producers of legal information sources, such as the Italian Parliament, various ministries, and a number of commercial publishers.

As concerns legislative information, i.e. statutory rules and regulations, the full text of statutes is offered free of charge in the Web site of the Parliament, the sites of various ministries and that of the Istituto Poligrafico della Zecca dello Stato (IPZS)¹ which gives access to the full text of statutes published only in the last sixty days.

A particular mention is to be made of the NIR, *Norme in Rete*² national project (Section 4), an e-government initiative coordinated by the Italian Authority for Information

¹ IPZS publishes and distributes the Official Collection of Regulatory Acts and the Gazzetta Ufficiale della Repubblica Italiana (Official Gazette of the Republic of Italy), which represents the well established, official collection of new laws, regulations, public work contracts and job offers. It

² Information on NIR Project are available on <http://www.nir.it>

Technology in the Public Administration in conjunction with the Ministry of Justice, ITTIG and other institutions. The project aims to improve accessibility to legislation by providing a unique point of access to Italian legislation published on different Web sites.

NIR is based on a federative approach through the cooperation among a number of providers of legislative information which adopt specific national standards and tools developed to facilitate their adoption (4).

As regards judicial cases, printed material is accessible through legal journals and collections of maxims (case abstract) of the decisions of the Italian Supreme Court. Online retrieval of case law is offered by the Italgire Legal Documentation System of the Supreme Court of Cassation (CED)³, which is a major Italian documentation source for legal information.

As regards legal literature, access to it is a primary requirement for understanding and interpretation of statutes and cases.

In Italy the difficulty of delimiting the scope of legal literature's sources makes its access problematic as compared to the other two legal information sources (5).

The production of Italian legal literature is mainly in the hands of commercial publishers, but bibliographic access tools are provided also by libraries producing indexes for certain journal articles⁴. To these, the national indexing service DoGi-Dottrina Giuridica⁵ must be added. As regards electronic legal literature, this is still rather scarce in Italy, whereas the production of legal electronic resources is nowadays quite widespread in a number of countries.

3. Linking legal sources

It is a matter of fact that today a vast and diversified amount of legal information is available worldwide. In this context it is of paramount importance to implement methodologies and tools for establishing context sensitive links in a way that, starting from legal literature's resources, users be directed to legislative information and case law reports.

In particular in civil law systems legal literature documents make reference to legislation, case law and also to additional legal literature material. In judicial case reports only statutory rules and other cases are cited, whereas laws make reference only to other laws. Fig.1 shows the combination of possible links among these legal sources.

³ From 1980 onwards, there was a gradual development and consolidation of the system, now offering access not only to case law, but to a number of legal literature and legislation resources. From late 2003 a new search interface is available, called Italgire Web (<http://www.italgiure.giustizia.it>). Information are available at <http://www.giustizia.it/cassazione/infoced.htm>

⁴ Catalogo delle pubblicazioni periodiche delle biblioteche aderenti a ESSPER. <http://www.biblio.liuc.it:8080/biblio/essper/period.htm>

⁵ DoGi is one of the most relevant sources for online research of legal literature in Italy. The database provides abstracts of articles published in more than 250 Italian periodicals. The database is available at <http://nir.ittig.cnr.it/dogiswish/IndexEng.htm>

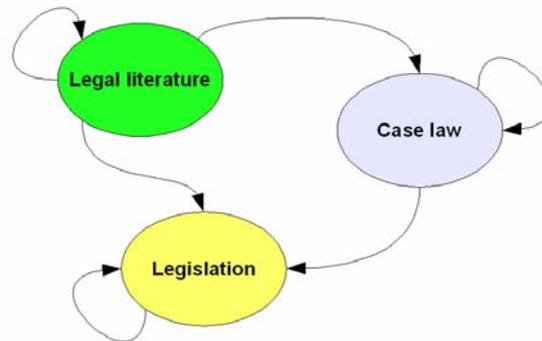


Fig. 1 The linking scenario in the legal sources domain

3.1 The need for intelligent identifiers

The enormous growth and dominance of the Internet and in particular the World Wide Web has made Uniform Resource Location (URL) almost synonymous with an identifier of an entity (or document). However, such identifier cannot be used as an effective reference point since it lacks stability and persistence.

References based on physical locations, expressed through URLs, presents the following well-known problems (6):

- difficulty in knowing the location of the cited resource;
- loss of validity over time of the document locations;
- impossibility of describing references to the resources not published yet.

To overcome these problems identifiers to electronic entities must be based not on the physical location, but on the logical content, thus on Uniform Resource Name (URN).

URNs were conceived for providing unique, unambiguous and lasting identifiers of network resources, independently of their physical locations.

Identifiers should be intelligent, built for ready interpretation outside the identifier scheme and able to derive meaningful information. An example of intelligent identifier is the Serial Item and Contribution Identifier (SICI) for a periodical's article, which contains sub-strings denoting elements such as the ISSN, the first characters of the title of the article, issue, volume, date of publication, page number, etc.

For the Portal under development the adoption of URNs identifying legislation, case law and legal literature has been considered a major requirement.

4. Identifiers for legislation

As discussed in Section 2, the NIR project has proposed the adoption of standards for legislative documents.

Two main standards have been defined: an XML standard for different kinds of legislative documents, formulated by three DTDs of increasing degree of complexity⁶, and a standard for legislative document identification, defined according to the URN technique⁷ (6) (4).

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

⁶ Specific DTDs for the NIR Project have been elaborated. F. Megale, F. Vitali. I DTD dei documenti di Norme in Rete. In *Informatica e Diritto*, 2001, n. 1, pp. 167–231

⁷ Established in conformity with those defined within IETF (Internet Engineering Task Force) RFC 2141 *URN Syntax* (7).

- a schema for assigning names capable of unambiguously representing any legal measure, issued by any authority at any time (past, present and future);
- a resolution mechanism from a uniform name to the on-line location of the corresponding resources.

The URN-NIR schema (6) includes a namespace identified by “`nir`” (this identifies the context in which the names are valid) and a syntax to represent the measure details:

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

The specific name `<NSS-nir>` must contain information to unambiguously identify a document. In a legislative environment, it is also necessary to distinguish among any later versions of a document and among its amendments. In this case information regarding a specific version of a document is added.

The general structure of a specific URN-NIR is therefore:

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

The `<document>` part is composed by document information related to the enacting authority, the type of measure, the details and any annex:

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

Examples of URNs of legislative documents are here reported below:

Act no. 675 of 31 December 1996

```
urn:nir:stato:legge:1996-12-31;675
```

Decree of Ministry of Finance of 20.12.99

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

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

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

URN-NIR is therefore independent from the availability of the document, from its physical location and its access mode. Only the significant details of the document and the knowledge of the URN syntax are necessary to the identification of the act.

Within the NIR domain, URN technique is used to describe cross-references, retrieve legal measures, represent relationships among legal documents.

The association between uniform names and documents can be obtained by using meta-information within the document (ex: in HTML files using a META element, in XML files using a suitable element) or maintained outside the documents but strictly related to it (by specific attribute in a database, or using methods as RDF technology).

In order that URNs be effective in hyperlinks, a resolution mechanism able to associate the URN of a document to one or more physical locations of it is necessary.

The URN technique allows also the development of parsing tools able to automatically detect references and construct the related URNs on the basis of the textual citation of the referred document and the knowledge of the URN syntax only.

Moreover, such techniques allows to construct a knowledge base storing the relations among legal documents.

4.1 NIR resolution service for legislative documents

The resolution service aims at associating the physical locations on the net to a uniform name. Two main components are necessary in a distributed environment: 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.

The resolution service is composed of a knowledge base or a set of transformation rules and a software module able to query the knowledge base. Through the NAPTR records of the DNS (described in RFC 2168 (8)), the client identifies the features (protocol, port, site) of the service able to associate a URN to different URLs, and, therefore, allowing to access to different copies of the same document, published by different data providers. Here below an example of the NIR resolution mechanism invocation is reported:

```
http://www.nir.it/cgi-  
bin/N2Ln?urn:nir:stato:legge:1996-12-31;675
```



Fig. 2 The results of a query, containing a URN-NIR, to the NIR resolver.

In Fig. 2 the results of the previous query to the NIR resolver is shown.

4.2 The catalogue for resolution

The implementation of a catalogue provides flexibility in resolving incomplete or incorrect URNs; this usually happens in legal domain as a result of a cataloguing activity but even more if the URNs have been automatically constructed (using a parser) from the textual wording of the references.

Apart from document URLs, the catalogue can contain a series of other information for the user, such as metadata (like title, details of publication, subject, in force date, alias, as well as the active and passive relations among other documents such as amendments, delegations, repeals, conversions, etc.).

4.3 The NIR resolution process

The NIR resolution process is composed of the following steps (see Fig. 3):

normalisation of the uniform name to be resolved, namely the transformation of some components to the canonical form (as normalising institution names, the type of measures, etc.);

querying the catalogue with normalised URNs and extracting the associated URLs (*resolution*), as well as any possible metadata.

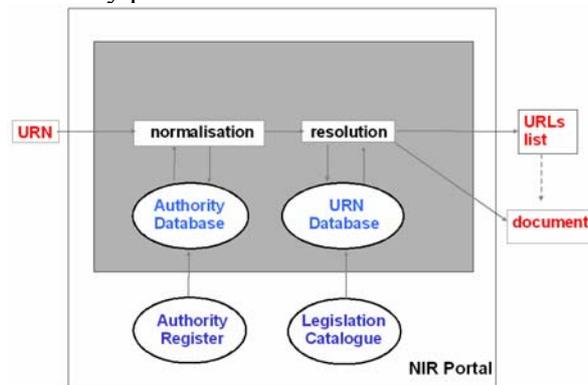


Fig. 3 The NIR resolution mechanism

5. Identifiers for case law

The NIR project has also addressed the problem of case law identification in a distributed environment. The definition of a URN scheme for case law documents is still under development; however a preliminary grammar has been proposed on the basis of that for legislation. The components are basically the same of the URN grammar for legislation (Section 4). Differences refer to the vocabulary to be used for:

- the enacting authority (<authority> field) that contains the normalized names of the Courts issuing a specific decision, as well as the types (such as civil, criminal, labour, etc.) and/or the number of their sections;
- the type of measure (<measure> field) containing normalized names of the types of decisions: court order, sentence and decision (this latter for Council of State);
- the details (<details> field) containing dates and numbers in a normalized form as well.

Examples of URNs for case law documents are:

Court order of the Court of Cassation, criminal section III n. 123 of 15 December 1997
`corte.cassazione;sezione.penale.3:ordinanza:1997-12-15;123`

As for legislation, a similar resolution mechanism will be built for case law URN scheme.

6. Development of standards for legal literature material

Having established standards for legislation and case law, the definition of a standard for legal literature identification is desirable, so to obtain a fully connected legal documents scenario (Section 3, Fig. 1), as well as to provide high quality retrieval services, within the Portal for legal literature under development (Section 1).

Identifiers available for bibliographic material in general are suitable for legal literature, since this type of literature has much in common with the rest of information resources included in the bibliographic domain.

But an intelligent and self-explanatory standard identifying the various types of documents in which legal literature's items appear, has not been developed so far. This is the reason why the legal community is highly interested in developing a community-legal literature URN scheme for coordinated access across repositories. In the Portal under

development, data providers are federated through a DC metadata scheme as a target bridging format to describe resources. Within a similar federative accordance, this URN standard can be effectively used to identify resources as content of the *dc:identifier* metatag.

There is a relevant number of bibliographic identifiers, some of which are very well established and widely used such as ISBN, ISSN, SICI, BICI. In particular, in the commercial environment the DOI – Digital Object Identifier⁸ is an important international standard for the identification of electronic information resources. However DOI and the other above mentioned identifiers are not adequate to express legal literature in a self explanatory way, because they are not context sensitive. For example DOIs suffer from the limitation that point users to the same articles, regardless of a user’s affiliation, missing the “appropriate copy” requirement.

Therefore a mechanism is proposed for the construction of a legal literature identifier, expressing a resource in a self explanatory way. It is based on data format standardization: producers of legal literature (authors, publishers, distributors) are free to manage their own information system, but all should agree on a common document identification technique and standard data scheme. A cooperative layer ensures the availability of common services and shared information.

This new standard reflects the features of legal literature material, including the elements usually adopted in bibliographic citations. As legal literature can appear in a variety of genres (books, chapters of books, journal articles, proceedings, theses, legal commentaries on cases and legislation, unstructured Web resources, etc.), and owing to the fact that the genre is not always directly expressed in bibliographic citations, the URN structure has been defined as independent from it: if available it can be manually inserted within a URN, providing more focused information able to facilitate the resolution process.

In order to identify the bibliographic elements of legal literature sources, the BibTex⁹ entries scheme used to compile bibliographies in the Latex framework has been considered (Tab. 1).

BibTex main entry types	Mandatory elements
Book	author or editor, title, publisher, year
article	author, title, journal, year
inproceedings	author, title, booktitle, year
inbook	author or editor, title, chapter and/or pages, publisher, year
Master Thesis/PhdThesis	author, title, school, year
proceedings	title, year
unpublished	author, title, note

Tab. 1 Bibtex main entries and their mandatory elements for the description of bibliographic resources.

On the basis of these main elements, a URN syntax to identify legal literature resources is proposed, adopting “lld” (acronym for *legal literature domain*) namespace:

<URN> ::= "urn:lld:" <NSS-1ld><NSS-1ld> ::= <resource>

⁸ <http://www.doi.org>

⁹ <http://www.ctan.org/tex-archive/biblio/bibtex/>

Each resource is described by its author/s or editor/s last name, its title, the publication date, optionally some details (including for example the parent title) and the genre.

```
<resource> ::= (<author-last-name>|<editor-last-name>) *["+" (<author-last-name>|<editor-last-name>)] ":" <resource-title> ":" <date>?[":" <details>] ?["@" <genre>]
```

These main elements are deeply expanded as follows:

```
<date> ::= <year>?[";" <month>] ? ["!" <edition-statement>]

<details> ::= <parent-title> ?["!" (<chapter-number> | <journal-number>)] ";"
<first-page-number>

<parent-title> ::= <parent-title-first-seven-words> ?["!" <publisher>]

<journal-number> ::= <issue-number> ?["$" <volume-number>]

<resource-title> ::= <title-seven-first-words> ?[";" <type-legal-contribution>]

<type-legal-contribution> ::= "report.on.case" | "comment.on.case" |
"report.on.statute" | "commentary.to.statute" | "book.review"

<genre> ::= ("book" | "article" | "inproceedings" | "inbook" | "master.thesis" |
"phd.thesis" | "proceedings" | "unpublished")

<year> ::= 4<digit>
<month> ::= 2<digit>
<edition-statement> ::= 1*<digit>
<parent-title-first-seven-words> ::= 1*<normal>
<first-page-number> ::= 1*<digit>
<chapter-number> ::= 1*<digit>
<issue-number> ::= 1*<digit>
<volume-number> ::= 1*<digit>
<title-seven-first-words> ::= 1*<normal>

<normal> ::= <lowercase>|<digit>|"."
<lowercase> ::= "a" | "b" | "c" | "d" | "e" | "f" | "g" | "h" | "i" |
"j" | "k" | "l" | "m" | "n" | "o" | "p" | "q" | "r" | "s" | "t" | "u" | "v" |
"w" | "x" | "y" | "z"
<digit> ::= "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"
```

This is an example of a URN for legal literature resource:

Sorge, C. and Bergfelder, M. (2004), Signatures by Electronic Agents: A Legal Perspective, In: Cevenini, C. (editor), *The Law and Electronic Agents: Proceedings of the LEA workshop*, Bologna, June 2004, pp. 141-153

```
urn:lld:sorge+bergfelder:signatures.by.electronic.agents.a.legal.perspective:2004;06:the.law.and.electronic.agents.proceeding.of;141@inproceedings
```

6.1 A resolution mechanism for legal literature

A resolution mechanism for legal literature identifiers will share the same principle of the resolution mechanism for legislation and case law identifiers.

The retrieval of resources is achieved in two different ways (2). In case of bibliographic metadata coming from structured repositories, the URNs of resources are firstly harvested from the data providers and mapped to the *dc:identifier* element using the OAI-PMH approach. To retrieve a specific resource, a rule-based resolution mechanism is invoked,

basically an http query containing the URN itself of the resource combined with the GetRecord verb of the OAI protocol. Here below a query example is shown:

```
http://www.[domainName].com/OaiDataProvider/servlet?verb=GetRecord&identifier=0ai:dogi.ittig.cnr.it/[URN]&metadataPrefix=oai_dc
```

In case of Web resources, URNs should be inserted within a document using a previously defined html metatag or xml element; at the service provider level a Web crawler is able to harvest resources collecting URNs, then associating the related URLs.

The service provider may also give advanced services for context-sensitive linking (9) across different types of legal resource (as access to different databases and to other resolvers, links to the full-text, to search engines, to local catalogues and their services (document delivery, etc)). This can be obtained using the OpenURL (10) mechanism (a NISO standard) for transporting metadata. Fig. 4 shows the functionalities of the Portal's OpenURL resolver: starting from a citation to a statute, described by an OpenURL containing the statute URN, the full text of the cited statute can be retrieved.

Moreover full or partial URNs of commentaries of the current statute can be constructed on the basis of related information in the current statute URN. Thus related commentaries can be retrieved from OAI commentary repositories. Finally, other related documents can be retrieved as well by querying a search engine through metadata transported by the current OpenURL.

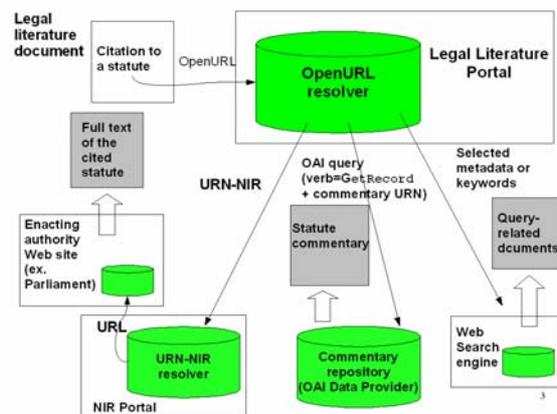


Fig. 4 The OpenURL + URN resolution mechanism

7. Conclusions

In the legal domain linking among different types of sources is of paramount importance. A stable management of the linking scenario among the three main legal sources (legal literature, case law and legislation) is desirable, so that legal experts can have a reliable access to the text of laws (legislation), to their interpretation (legal literature), and to the enforcement of these laws in judicial cases (case law).

The relation among different legal sources can be achieved through a mechanism based on their identification following a consistent URN technology.

In a distributed environment the use of such standards by data providers, together with the development of specific resolution mechanisms from URNs to URLs, allows to express cross-references in a stable way, independently from the physical location of the

referred documents. The proposed advanced services, based on an OpenURL resolver, can improve the quality and the effectiveness of information services for legal users.

References

1. D. Barmakian. Better search engines for law. In *Law Library Journal*, 2000, vol. 92, n. 4, pp. 399-429
2. E. Francesconi and G. Peruginelli. Access to Italian legal literature: integration between structured repositories and Web documents. In *Proc. of the DC Conference 2003*, pp. 99-107
3. E. Francesconi and G. Peruginelli. Retrieval of Italian legal literature: a case of semantic search using legal vocabulary. In *Proc. of the DC Conference 2005*, pp. 97-105.
4. C. Biagioli, E. Francesconi, P. Spinosa, M. Taddei, Legislative Drafting Support Tool Based on XML Standards. In *Proceedings of the International Conference on E-Government, 2005*, pp. 390-397
5. G. Peruginelli, L'accesso alla dottrina giuridica: strumenti e linee di sviluppo in rete. In *Informatica e diritto*, 2002, vol. XXVIII, n. 1, pp.111-176
6. P. Spinosa, Identification of Legal Documents through URNs (Uniform Resource Names). In *Proceedings of the EuroWeb 2001, The Web in Public Administration*.
7. R. Moats, K. R. Sollins, *URN Syntax*, RFC 2141, May 1997.
8. M. Mealling, R. Daniel, *The Naming Authority Pointer (NAPTR) DNS Resource Record*, RFC 2915, September 2000.
9. A. Apps, R. MacIntyre, Using the OpenURL Framework to Locate Bibliographic Resources. In *Proc. of the Dublin Core Conference 2003*, pp.143-152.
10. H. Van de Sompel, Open Linking in the Scholarly Information Environment Using the OpenURL Framework. In *D-Lib Magazine*, 2001, vol. 7, n. 3. <http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>