BUILDING AN INTEGRATED WEB SERVICE TO ACCESS ITALIAN LEGAL LITERATURE

A cura di
Ginevra Peruginelli

Rapporto tecnico n. 10/2003
Firenze
INDEX:

Executive summary p. 1

1. Introduction p. 1

2. Overview of Italian situation: services and user needs p. 2

3. Objectives of the project p. 3

4. Project planning and development p. 3
   4.1 Selection of information content p. 3
   4.2 System development and technical components p. 4
   4.3 Management, organisational aspects and evaluation of the system p. 5

References p. 7
Executive summary

This project is a case study focusing on legal literature’s search and retrieval, attempting to harmonise access to different data sources providing information on Italian legal doctrine. It is proposed as a free web service offered by ITTIG, Istituto di Teoria e Techniche dell’Informazione Giuridica of CNR\(^1\), in line with its institutional mission to support users in legal research. After a brief analysis of available services in Italy and legal doctrine users’ requirements, planning and implementation of the new service are described. The first stage of the project concerns the selection of material. For the specific prototype project three types of data sources are chosen: a limited number of OPACs, a journal article data base, and a publisher’s on line catalogue. The second phase includes the definition of functionality to be offered: essentially resource discovery and retrieval, while access to documents is left to the local systems users are addressed to. Technical aspects’ definition follow, including the adoption of the protocol for integrating resources in a consistent way (OAI-MHP), managing interoperability issues, mapping preparation between different metadata and adoption of Dublin Core as bridging format, comparison between subject indexing access points, interface design. Organisational and management issues relate to preparatory and ongoing activities, where a major role is played by the service provider, ITTIG, in allocating resources of various type: staff, technical facilities, funds for upgrading its software and hardware, if needed. Of particular importance is its commitment in promoting and maintaining the service and in convincing and supporting data providers to adopt standard metadata, opening up their repositories to worldwide access. Evaluation of the system by a representative user sample is planned, to orient the development of the system on a technical and functional basis. Complex issues are the involvement of heterogeneous data providers, mapping between metadata, comparison between different classification systems, design and management of the system. Functionality such as profiling, current awareness, seamless access to print based documents and primary web resources are to be scheduled in a further stage of the project. By this project ITTIG places itself in a position that fosters co-operation with legal literature’s data providers, while offering legal users enhanced facilities to better exploit the resources they need in their research work.

1. Introduction

This project focuses on integrated access to Italian legal literature. Legal doctrine, together with statutes and case law, 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 students and scholars, judges and lawyers, administrators, politicians, reference librarians and information professionals, journalists, ordinary citizens. More and more, under the pressure of globalisation and technology which both allow for massive and rapid production and delivery of information content,
users demand improved, harmonised and fast access to a wide range of legal literature resources, asking at the same time for quality control. The Institute (ITTIG) I work for seems the right candidate organisation for supporting a project meeting these needs, for its long-standing activity in the legal research domain. The Institute’s focus is both on research projects (such as NIR, Norme in Rete [1], aimed at providing a unique point of access to Italian statutes in digital form) and on development of information access services, such as the production of DoGi – Dottrina Giuridica^2, a bibliographic database including TOCs and abstracts of articles of main Italian legal periodicals. The development and management of the proposed project require new skills by its staff, new tasks and responsibilities, a strong commitment, co-operation ventures and decision-making. Before describing the project, a brief analysis is presented of Italian available tools and services in this area, followed by an outline of user requirements.

2. Overview of Italian situation: services and user needs

Primary and secondary sources of legal literature are found in different locations, institutions, sites and systems, showing a wide variety in terms of:

♦ type of documents produced: journal articles, monographs, essays and reports, opinions on case law, working papers, theses, grey literature; they are mostly on paper, but electronic resources are starting to be produced and made available on the web;
♦ bibliographic and indexing tools for information retrieval (OPACs, indexing services);
♦ data representation (metadata formats and their encoding conventions);
♦ access interfaces;
♦ conditions of access and use.

In Italy, as abroad, legal users extensively rely on OPACs and bibliographic tools for searching, but as compared to the international landscape, differences are found in the nature and offerings of indexing services and in legal electronic resources’ production. The production of Italian legal literature is mainly in the hands of commercial publishers, and bibliographic access tools are provided by libraries through their OPACs, by groups of academic libraries producing indexes for certain journal articles^3, and by a few specialised publishers. To these, the national indexing service DoGi must be added. All these services have various interfaces and adopt different classification systems, causing troubles to legal users. Abroad the production of legal literature is, in general, institutionally based, mainly published by universities. Mostly used access tools are OPACs that point to collections of monographs and periodicals (including remote e-journals), whereas indexing services are commercially based, accessible on subscription

^2 DoGi is accessible on the web, free of charge, and it represents, despite some deficiencies in coverage and currency, the most important source of articles of Italian legal printed periodicals. [http://www.ittig.cnr.it/BancheDatiGuide/dogi/Index.htm](http://www.ittig.cnr.it/BancheDatiGuide/dogi/Index.htm)

by users of law libraries and external users who can afford to pay high rates for such services⁴.

As regards electronic legal literature’s production, it is still rather scarce in Italy, whereas internationally legal electronic resources are quite widespread. Legal literature users are a very heterogeneous group [2], from novice to very expert and demanding [3]. In Italy not all of them are completely at ease using OPACs and web based electronic indexing services. Known item and topical searching are generally both highly required, depending on the level of research, but different search interfaces disorientate and sometimes even discourage a good part of them to use electronic access services. A certain part of scholars and professionals simply rely on browsing their preferred printed reviews and directories and some almost ignore the existence of automated tools for searching.

All this implies that access tools and services have to carefully take into account the requirements, computer skills and attitudes of these users, trying to reach potential users, offering information assembled in a consistent way, without constraining them to search multiple sources and different systems. An added value functionality to be offered is to address users, through pertinent links, from legal literature references to related statutes and case law reports.

3. Objectives of the project

Based on such situation, the proposed project attempts to build a web-based access service to Italian legal literature by integrating, at its start, a limited number of data sources which provide information on legal literature.

The objectives of the project are:

a) Improving access to Italian legal literature by setting up a user centred service providing resource discovery and location of distributed and heterogeneous resources, offering a common user interface and an integrated view

b) Generating increased confidence of legal literature users in technologies supporting automated searching and in integrated services⁵

c) Devising and experimenting methods and tools to enhance access to differently classified resources by exploiting the semantics of subject metadata prepared by different organisations

d) Initiating a close co-operation between data and service providers in legal literature’s field and facilitating shared acquisition of competence and attitudes towards standard solutions.

⁴The most important and highly used services are Lexis-Nexis and Westlaw, which are world wide, on charge accessible databases containing TOCs, abstracts, and digital texts, whose size is growing progressively.

⁵This purpose is based on the importance of communication and information technologies in learning, teaching and research activities, as stated in IMPEL2 Project.

http://online.northumbria.ac.uk/faculties/art/information_studies/impel/back.htm
4. Project planning and development

The project requires that the following tasks be taken into account, while coping with constraints such as staff resources, funds and time.

♦ Selection of information content
♦ System development and technical components
♦ Management and organisational issues for the setting up and running of the service, including measures for the evaluation of the system.

4.1 Selection of information content

In implementing this type of integrated service it is essential to establish a strategy for the selection of both material and producers. Users need reliable resources in terms of quality of content and their description, currency and coverage of information, authority of producers, accessibility (resources can be free or for a fee, both requiring clear indication of subscription and registration).

Selection of material is an ongoing process and essential activities are the updating of the system with new resources and the periodical monitoring of the existing ones, thus achieving currency and also consistency of metadata. Users need to know what can be done with retrieved material (ILL, document delivery, view, download). Therefore, when selecting accessible resources for inclusion into the system, conditions of use must be checked, and helpful guidelines prepared in the system’s home page, together with a description of accessible data sources. In this project access to documents is left to the facilities of each selected service, where users can locate the retrieved materials and put requests.

It seems feasible to start the project by selecting a limited set of resources which should be representative of important sources of legal literature in order to generate interest in the service from its start. The system will include DoGi\(^6\), the main bibliographic data base of Italian legal review articles, a number of OPACs\(^7\) (from which relevant law material records are to be extracted by using Dewey Decimal Classification 34X notation), and a publisher’s catalogue\(^8\). This mix should allow for the creation of a reliable test bed to evaluate the choice made of accessible resources, users’ reaction and, most important, the technical and management issues associated with the setting up of the system.

4.2 System development and technical components

The technical components include the design of the new system and the adoption of the integration model, hardware and software facilities, preparation

\(^6\) DoGi provides, together with TOCs and rich abstracts, links from literature to legislation and case law sources. It also offers a link to periodical titles described in the National Union Catalogue of periodicals (ACNP: Accesso al Catalogo Nazionale dei Periodici. URL: http://acnp.cib.unibo.it/cgi-ser/start/it/cnr/fp.html) where libraries’ holdings are visible.

\(^7\) OPACs of Florence University (http://opac.unifi.it/), of Biblioteca Centrale Giuridica (http://opac.giustizia.it/sebina/opac/ase) and of National Central Library of Florence (http://catalogo.bncf.firenze.sbn.it/cgi-opac/opac.cgi).

\(^8\) Giuffrè Editore (http://www.giuffre.it/servlet/page?_pageid=60&_dad=portal30&_schema=PORTAL30).
of mappings between metadata formats used for the description of resources, solutions to semantic and technical interoperability problems, interface design and system development.

The choice between possible levels of integration in accessing distributed resources is the object of lots of studies and speculations [4] [5] [6]. It depends on several factors, such as the type of information resources and their description, technical constraints, available funds, conditions made by producers.

A low level of integration, widely adopted by many institutional web sites, libraries and information centres, is based on the ability of the world wide web to present heterogeneous information services to users through a simple menu interface, showing links to local and remote resources, where the role of the developer is to select resources to list and maintain links to on-line services.

An enhanced integration model does not limit itself to provide links, but concentrates upon a hybrid and distributed environment, aiming at building a framework for presenting users with a range of heterogeneous information services in a consistent and integrated way, such as that recommended in UK by the MODELS initiative: this is the model proposed for our project.

Z39.50 protocol and the OAI-MHP (Open Archive Initiative – Metadata Harvesting Protocol) have been considered for accessing different distributed databases.

As regards protocols Z39.50, quite popular and successful for OPACs cross searching, is rather complex to implement [7] [8]. It is subject to different interpretation even in a library context, and unlikely to be implemented in the servers of indexing services and commercial repositories that the project intends to reach. Some software tools have tried to emulate a Z39.50 server , but management is rather cumbersome.

The OAI approach is simpler to implement in our environment, where references to legal literature are found in an increasing number of sites and services, using different metadata conventions and formats. This implies the need to rely on a simple protocol to harvest their metadata to create added value access services for users. Although OAI has its origins in the e-print arena to facilitate rapid and open access to digital original material, it is now adopted in multiple contexts. Its principles are based on the distinction of two types of stakeholders: data providers, producing metadata of resources (their records being “OAI compliant”), and service providers, which harvest these metadata through the protocol (OAI-MHP) and implement added value services by exploiting the collected externally produced metadata. A main

9 MOving to Distributed Environments for Library Services (MODELS) is a UKOLN initiative supported by the Electronic Libraries Programme and the British Library. http://www.ukoln.ac.uk/models/


12 One example of implementation is the provision of ARCA server toolkit, developed for the realisation of the metaopac in Tuscany, Italy. ARCA Z39.50 tool kit. http://www.codices.com/arca.htm

13 See for example: Open Archives Initiative Virginia Tech DLRL Projects http://www.dlib.vt.edu/projects/OAI/#projects; University of Glasgow DEDALUS project. http://www.lib.gla.ac.uk/daedalus/links/
recommendation is that data providers deliver (this is done mainly by conversion) at least metadata corresponding to Dublin Core\textsuperscript{14} metadata element set, a standard minimal set of 15 elements describing their resources. Other points in favour are that several open source software tools exist for the implementation of the protocol and a rich documentation on projects and issues involved is available\textsuperscript{15}.

Another specific feature of OAI is its flexibility and feasibility to embrace (based as it is on XML - Extensible Mark-up Language), in the case of legal information, records relating to statutes and case law reports, thus facilitating integration between these two components and legal literature records.

What eventually has to be pointed out is that so far some barriers exist between library world and indexing services in using standards for describing and supporting access to the item as a whole and those to describe periodical articles. This is a barrier to integration, but relying on a limited set of elements (such as Dublin Core) functioning as common denominator or “lingua franca”, seems the right solution for combining these types of differently described resources for the sake of interoperability between various metadata.

The proposed project intends to experiment such approach for the data sources mentioned above, while recognising that the following technical issues are likely to be encountered:

- handling rich metadata as compared to Dublin Core, trying to exploit such richness for enhanced searching by users
- providing metadata normalisation, that includes several delicate activities: a) conversion of non- “OAI compliant” records (generation of XML and OAI record coding to allow proper processing of metadata, selected harvesting and updating); b) mapping between MARC and proprietary formats to Dublin Core element set; c) semantic comparison of different classification systems used in the different repositories’ records. This is a major task as the project intends to achieve a high level of semantic integration (at least on broad law categories and related subdivisions), offering users subject-classification access in a seamless way beyond differences in the original systems’ indexing practices.

In developing the system and designing its interface, provision has to be offered for working with different platforms and browsers. Care should be put in defining the “right” number of search options, without overwhelming users with too many choices. Browse options should also be offered, presenting users a list of subject terms to help them in searching. Classification, as noted above, is a crucial search parameter by which users are expected to put queries on legal literature, especially for an initial exploration on a specific branch of law.

4.3 Management, organisational aspects and evaluation of the system

Management and organisational aspects pertain to both the initial phase of the project and to ongoing activities.

Preparation activities include a detailed definition of project’s characteristics and phases, its approval by ITTIG, and allocation of resources in terms of staff

\textsuperscript{14} Dublin Core Element Set, version 1.1. \url{http://Dublincore.org/documents/dces/}

\textsuperscript{15} OAI related utilities. \url{http://www.openarchives.org/tools/tools.html}; Digital libraries: metadata resources \url{http://www.ifla.org/II/metadata.htm}
and funds. The need for special funds to be allocated must be verified, mainly for upgrading some hardware and telecommunication components and for software tools.

It is essential to set up and train a team in charge of the project, with different specialisation, such as technicians and information professionals expert in user services building, indexing and metadata matters.

Their role is to:

♦ validate the proposed integration model
♦ analyse metadata created by content producers and prepare metadata mappings, while checking the work done in this field at international level [10] [11] and examining available tools
♦ thoroughly examine integration models, OAI specifications, ongoing international projects in this area
♦ analyse the different indexing systems used by data providers and establish correspondence between classification metadata to allow for cross subject searching (this is likely to be a recurring process during the project’s life)
♦ investigate availability of software (open source and non).

ITTIG should commit itself to co-operate with data providers, persuading them to use standard protocols and common conventions for preparing and exposing their data.

Marketing and promotion of the new system is another important activity, to be done through the Institute’s web site and through advertising in legal reviews, in official legal web sites, law libraries and law firms offices. Training and help modules are to be scheduled.

Running activities include the regular maintenance of hardware and software components of the system and periodical analysis of system performance, its availability and reliability, security of access and data. Monitoring and evaluating the system is an important activity to be undertaken once the service is implemented and tested from the strictly technical point of view. In order to test its usability a plan has to be prepared, specifying the user sample and the evaluation methodology. It is proposed to select different categories of users, reach them by electronic and ordinary mail, telephone, informing them on the novelty and importance of the service. Gathering of results can be done through an on line form, including specific questions to measure satisfaction, through direct interviews and observation. Detailed analysis of user feedback should follow, in order to orient the system according to requirements in terms of user interface and overall functionality.

The project’s main concept, focusing on the provision of integrated access to legal literature bibliographic sources, is not revolutionary, but what seems important is the adoption, in the legal literature domain, of an open approach by different stakeholders for the benefit of legal users in their cross searching activities. A changing perspective of an institution such as ITTIG is also an important element, for its commitment in the setting up and management of a service which combines externally produced remote data sources.

16 See for example: Metadata: UKOLN software tools. http://www.ukoln.ac.uk/metadata/software-tools/
REFERENCES

[1] Studio di fattibilità per la realizzazione del progetto Norme In Rete. *Informatica e diritto*, n. 1, 2000 (numero speciale)


  [http://www.loc.gov/catdir/bibcontrol/tdg7.html](http://www.loc.gov/catdir/bibcontrol/tdg7.html)

[5] Liu, X. Federating heterogeneous digital libraries by metadata harvesting. A dissertation submitted to the Faculty of Old Dominion University, December 2002


  [http://www.ukoln.ac.uk/metadata/interoperability/](http://www.ukoln.ac.uk/metadata/interoperability/)


Web addresses checking date: 12/11/2003.