

PROVIDING ACCESS TO THE INACCESSIBLE: A NEW THEMATIC DIGITAL LIBRARY FOR THE VENICE LAGOON

Stefania Biagioni ⁽¹⁾, Erica Busatto ⁽²⁾, Carlo Carlesi ⁽¹⁾, Lisa Locatello ⁽³⁾,
Monica Ortolan ⁽²⁾, Antonella Zane ⁽²⁾

⁽¹⁾ "A. Faedo" Institute of Information Science and Technologies (ISTI-CNR) – Via G. Moruzzi, 1 56124 Pisa (Italy)
stefania.biagioni@isti.cnr.it, carlo.carlesi@isti.cnr.it

⁽²⁾ Vallisneri Library - University of Padua – Viale G. Colombo, 3 – 35131 Padua (Italy) –
ericabusatto@libero.it, monica.ortolan@unipd.it, antonella.zane@unipd.it

⁽³⁾ Department of Biology - University of Padua – Viale G. Colombo, 3 – 35131 Padua (Italy) –
lisa.locatello@unipd.it

Abstract

This paper is a description of an innovative project: the creation of a digital thematic library known as the *Laguna di Venezia Digital Library (LVDL)*. It will collect published and unpublished scientific documentation about the Venice Lagoon. LVDL is a personalised extension of the ETRDL (ERCIM Technical Reference Digital Library) system, which is based on the Dienst protocol. Thanks to varying levels of document search and semantic descriptors, LVDL will satisfy many kinds of users working in a wide range of topics. As organised, this digital library has the potential to become a key thematic library for multidisciplinary documentation on the Venice Lagoon.

Keywords: thematic digital library, Venice Lagoon, digital library prototype

Introduction

The Venice Lagoon is unique. No other place in the world has its particular historical, archaeological and environmental significance. Over the centuries, the Lagoon has undergone major changes. Natural changes, such as variations in sea level and fluvial sedimentary deposits. And changes resulting from human activity, such as alternations in the coastline brought about by lagoon fish-farming, industrial production and pollution, with significant impact on the balance in the lagoon ecosystem which, since 1987, has been, along with Venice itself, a World Heritage Site.

Only in recent years have scientific projects, with encouragement and funding from national and international agencies, begun to focus on the protection of this environment. There have been analyses of the water (fresh, brackish and marine), studies of the sediment, the lagoon flora and fauna, etc.

This work has resulted in a large volume of documents presenting the results of academic/scientific research, public- and private studies on environmental impact assessment (EIA) and the scientific activities of naturalist associations.

This material, while of great interest to the Italian and international scientific communities involved in the preservation and protection of coastal environments is, however, difficult to find by means of the more common bibliographic tools (catalogues and bibliographic citation indexes).

Many of these documents are difficult to trace either because they are grey literature publications (theses, reports, conference proceedings, pre-prints, etc.) or because they are papers published in Italian journals not indexed in international bibliographic databases.

The solutions adopted up to the present time to bridge this information gap include bibliographies and collections of print published works related to a specific subject, place, author or Institution, known as “miscellany”.

The Vallisneri Bio-Medical Library of the University of Padua houses various lagoon-related miscellany collecting print publications on research carried out by Academy’s faculty and researchers between 1930 and 1970. These studies are an invaluable reference for scientists studying the evolution of the lagoon environment. Since many of these publications are not included in any electronic or print catalogues or bibliographic databases and are not available in digitised full-text versions, they are potentially exploitable but actually inaccessible scientific data.

Recently, a number of Web sites published by Italian institutions [7, 9, 14] involved in the protection of the Venice Lagoon environment, have posted related lists of references and of full text documents. Thanks to their well-structured hyperlink framework and very simple search form, they

provide a valuable service in resource discovery but they are ineffective in information retrieval.

The major problems faced by this user community can be summarised as follows:

- scattered bibliographic sources
- the need to carry out systematic, structured search sessions using specific keywords and descriptors.

Project purpose and implementation of the archive

On the basis of these considerations and on information-needs analysis, the Vallisneri Bio-Medical Library developed, in collaboration with the Institute of Information Science and Technologies (ISTI-CNR, Pisa), a project aimed to set up a digital thematic library. The purpose was to facilitate the search for and access to this documentation by a wider, more or less specialised (local and remote, institutional and otherwise) user basin by offering an advanced on-line search-and-browse service based on international standards for document metadata and semantic description (classification scheme and thesaurus).

The opportunity arose at the end of 2001, when the Vallisneri Library joined the ETRDL (ERIC Technical Reference Digital Library) [1, 2, 3, 4, 8] and was hosted in the ISTI multidisciplinary experimental version.

ETRDL, set up in 1997 as a specialised sub-collection of the U.S. *NCSTRL (Networked Computer Science Technical Reference Library)* [12] is a distributed digital library which publishes the technical documents of the European Consortium for Informatics and Mathematics (ERCIM).

ETRDL is a self-archiving system based on RFC 1807 metadata (mapping in Dublin Core) and on the Dienst protocol [10]. It offers extended functionality of multilingual information access, search-and-browse and submission, withdrawal and administration services.

In order to satisfy the specific needs of our disciplinary field, we set up the "Laguna di Venezia Digital Library" (LVDL) [11] as a personalised extension of ETRDL. The customisation required accurate selection and planning of the repository features (record fields, file formats, languages) and policies (publication-date span, document type, classification and subjects). The project was co-financed by Centro di Ateneo per le Biblioteche (the Central Office of the Library System) and the Marine Biology Course of the University of Padua.

The result is a digital-library prototype, currently hosted on a CNR server in Pisa, which provides bibliographic, full-text, factual and multimedia information through personalised search-and-browse user interfaces. The library staff manages both the submission of new documents and the administration of LVDL collections through a specific Web interface.

In this early phase, the project is focused on material relating to marine biology (in digital, digitised and/or paper format)

preserved at the following Institutions of the University of Padua: the Vallisneri Bio-Medical Library (Padua), the Marine Biology Course (Palazzo Grassi, Chioggia) and the Hydrobiological Station (Chioggia). At the present time, the archive includes about 150 documents, mainly theses, conference proceedings and published papers from 1930 to date.

In the LVDL, each institution is represented by a separate collection; it is also the location from which the user can obtain a print copy of all the documents included in the repository.

System functionality

1- Multilingual access

One of the most interesting features of the system is the cross-language querying functionality, which operates as follows:

- the title is always provided both in the original language and in English
- the use of keywords, descriptors and classification codes is always in English, making it possible to retrieve documents in other languages as well
- the abstract is provided in Italian (mandatory) and in the original language; the Italian papers' abstracts are always accompanied by an English version.

2- Submission and administration services

By means of the document submission form, each document is described with a set of metadata (RFC 1807)

- *titles* in original language, translated into English;
- *author*;
- *e-mail* address of the administrator lvdl@bio.unipd.it ;
- *subject*: three levels of semantic descriptors are provided:
 1. ASFA (Aquatic Sciences and Fisheries Abstracts) Thesaurus;
 2. LC (Library of Congress) classification;
 3. free keywords;

the LC Classification outline and the ASFA Thesaurus are accessible online, enabling the users to select the most appropriate codes and terms;

- *collection*: the Vallisneri Library (Padua), the Marine Biology Course (Palazzo Grassi, Chioggia) and the Hydrobiological Station (Chioggia) (by selecting the option from an adjacent drop-down menu);
- *abstract* in Italian and in the document's original language;
- *bibliographic description*;
- *publication dates*;
- *type of document* recorded, such as: technical reports, proceedings, maps, data sets, pre-print formats, graduate or undergraduate theses, published documents,

photographs, bibliography (selected from the drop-down menu);

- document *language* (selected from the drop-down menu).

Whenever possible, the *full text* of the documents is provided. The system supports various file formats (PDF, Ms Word, ASCII text, HTML, JPEG, PostScript, TIFF, GIF).

An administration interface enables the librarian to verify the state of the data and of the document contents before the documents are entered in the appropriate collection in the digital library. In addition, the administrator has the option of removing a document from the collection.

3- Search-and-browse services

Users begin the search by clicking on the Search/Browse options on the system homepage. Options for browsing the collection according to author, year and semantic indexes are available for each collection.

Users can search for documents (Fig. 1) using either a:

- *simple search*, consisting of a single field, in which the query is entered. The search returns all documents which contain the keywords entered in the field author, title, abstract and subject;
- *field search*, enables search in one or more of the following fields:
- the bibliographic fields: title, author, abstract in other language, subject
- three selectors to search documents according to type, year, language
- a menu for selecting one or more collections in which to perform the search and a check box to select all collections;
- *direct search*, consisting of a text-entry field in which a unique document identifier can be entered to access a specific document directly.

Future developments and aims

Excellent response from users to the first results of this initiative have encouraged us to continue to develop the project, taking into consideration new opportunities relative to both contents and technology:

- porting to the new, OAI compliant, Digital Library Management System “Open DLib” [5, 6, 13]
- integration with the upcoming University of Padua “Digital Library” Project
- activation of the self-archiving/publishing functionality, to be evaluated on the basis of requests from authors
- development of the collection in a multidisciplinary framework
- promotion of new collaborations with public and private bodies in step with increase in the number of collections
- development of the current LVDL homepage into a portal giving access to resources and Web sites related to the lagoon environment.

The image shows a screenshot of the LVDL search form. At the top, there is a logo for 'Laguna di Venezia Digital Library'. Below the logo, there are three main search sections: 'Simple Search', 'Fielded Search', and 'Direct Search'. 'Simple Search' has a single input field for 'Term(s)'. 'Fielded Search' has multiple input fields for 'Title', 'Author(s)', 'Abstract', and 'Subject(s)'. It also includes a dropdown for 'Other language' and a section for selecting collections. 'Direct Search' has an input field for 'Document ID'. The page also includes a 'Logical operator between fields' section and a 'Refine search' section with dropdowns for 'Type', 'Year', and 'Language'. At the bottom, there is contact information and logos for 'LVDL' and 'etrdl'.

Fig. 1: LVDL search form

Conclusions

LVDL is unique because of the originality of its thematic content. Thanks to a threefold level of semantic descriptors (free keywords, ASFA thesaurus, LC classification), it is able to satisfy the search and information-retrieval needs of many kinds of users, more or less specialised in biological and naturalist topics. In addition, the multilingual access functionality, which enables the search for and retrieval of documents written in wide range of languages, facilitates the visibility of project and research results and the dissemination of those results among scientific communities. Supported by its collaboration with public and private institutions involved in scientific studies on the Venice Lagoon, LVDL is intended to become a major multidisciplinary, thematic digital library. Within this context, the LVDL system may be considered as a prototype for DLs in the field of environmental sciences.

References

- [1] A. Andreoni, M.B. Baldacci, S. Biagioni, C. Carlesi, D. Castelli, P. Pagano, C. Peters. “*Developing a European Technical Reference Digital Library*”. Research and Advanced Technology for Digital Libraries. 3rd European Conference, ECDL 1999. Proceedings (Paris, 1999), 343-362. Lecture Notes in Computer Science 1696. S. Abiteboul and A.M. Vercoustre (eds.), Springer (1999)
- [2] A. Andreoni, M.B. Baldacci, S. Biagioni, C. Carlesi, D. Castelli, P. Pagano, C. Peters, S. Pisani. “The ERCIM Technical Reference Digital Library: Meeting the Requirements of a European Community within an International Federation”, D-Lib Magazine, 5, n. 12 (1999).
- [3] M.B. Baldacci., S. Biagioni , C. Carlesi, D. Castelli, C. Peters. “Implementing the Common User Interface for a Digital Library: the ETRDL Experience”, 8th Delos Workshop on User Interfaces for Digital Libraries. ERCIM Workshop reports. (Kista, Sweden, 1998)
- [4] S. Biagioni, J. Borbinha, R. Ferber, P. Hansen, L. Kovacs, F. Roos, A.M. Vercoustre - *The ERCIM Technical Reference DigitalLibrary*. Research and Advanced Technology for the Digital Library: 2nd European Conference, ECDL 1998. Proceedings (Heraklion, Crete, Greece, 1998), 905-906. Lecture Notes in Computer Science 1513. C. Nikolaou, C. Stephanidis (eds.), Springer (1998)
- [5] D. Castelli, P. Pagano. “OpenDLib: a Digital Library Service System”. In: M. Agosti and C. Thanos, editors, Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2002), pages 292-308. Lecture Notes in Computer Sciences (LNCS) 2458, Springer, Berlin (2002) .
- [6] D. Castelli P. Pagano. “*A System for Building Expandable Digital Libraries*”, Joint ACM-IEEE Conference on Digital Libraries (Houston, Texas, May 2003). Proceedings, p. 335-345. Catherine C. Marshall, Geneva Henry and Lois Delcambre (eds.). IEEE (2003)
- [7] CORILA Consorzio per la Gestione del Centro di Coordinamento delle Attività di Ricerca inerenti il Sistema Lagunare di Venezia <http://www.corila.it/>
- [8] ERCIM Technical Reference Digital Library <http://dienst.iei.pi.cnr.it/>
- [9] Forum per la Laguna <http://www.forumlagunavenezia.org>
- [10] C. Lagoze, J.R. Davis. “Dienst: an Architecture for Distributed Document Libraries”, Communications of the ACM, 38 n. 4, p.45 (1995).
- [11] Laguna di Venezia Digital Library <http://archimede.isti.cnr.it:81/introen.htm>
- [12] Network Computer Science Technical Report Library. <http://www.ncstrl.org>
- [13] OpenDLib <http://www.opendlib.com>
- [14] SAL.VE <http://www.salve.it>