The ERCIM Technical Reference Digital Library

A service and a testbed for the IT Research Community

The ERCIM Technical Reference Digital Library

(ETRDL)

The aim of the ERCIM Technical Reference Digital Library (ETRDL) is to assist ERCIM scientists to make their research results immediately available world-wide, and to provide them with appropriate on-line facilities to access the technical documentation of others working in the same field.

The ETRDL collection consists of all kinds of grey literature produced by the ERCIM Institutions (technical reports, proceedings of conferences or workshops, theses, project deliverables, etc.) and is managed by a set of interoperating servers. Pilot server sites have been set up at seven of the 14 ERCIM national labs. Servers are expected to be installed in other centres soon.

The system has been based on the Dienst system developed by a US consortium led by Cornell University and adopted by NCSTRL (Networked Computer Science Technical Reference Library). However, additional functionality is implemented in the ETRDL service in order to meet the particular needs of the European IT scientific community. ETRDL can also be used as a test-bed for ERCIM scientists working on DL related research issues.

The intention of this handout is to give the reader an idea of the system functionality and the Common User Interfaces (for Document Search and Submission, and for Administration). A description is also given of advanced functionality experimented by partners in the ETRDL project.

The system is publicly available at the DELOS Web site:

http://www.iei.pi.cnr.it/DELOS/ETRDL

The ETRDL Service

The ERCIM collection is part of the NCSTRL collection. A user accessing the NCSTRL service can view and query any of the collections of the ERCIM institutions using the standard NCSTRL search functions. The ETRDL system represents an extension of the standard Dienst system as a set of additional browse and search functions and an extended metadata set has been added. The new functionality has been implemented in order to meet the specific requirements of the European scientific community.

The ETRDL Users

Our technical reference digital library has been designed to meet the needs of three distinct classes of users within the ERCIM community:

- information seekers, i.e. people who will access one or more of the document collections available to find pertinent material;

- information providers, i.e. authors, or their representatives, who will submit new documents to a specific collection with associated bibliographic records;

- information administrators, i.e. those responsible (usually, but not always, the librarians) for verifying the correctness of the bibliographic records and the associated document files before inserting them into the appropriate collection.

The main system functionality had therefore to cater for the needs of these very different types of users: access/searching for information, submission/elimination of information, management of information.

The Metadata Set

In the ETRDL collection each document has a common metadata description associated. This description is based on the Dublin Core metadescription standard and represents an extension of the basic Dienst metadata set.

The user can employ the ACM Computing Classification and/or the AMS Mathematics Subject Classification, and/or free keywords to represent subject terms for document classification during the submission procedure and for retrieval when querying the system. The ACM and AMS schemes are accessible on-line and can be browsed during both retrieval and submission; codes with associated descriptors can be selected and inserted in the appropriate fields. Authors must enter codes/descriptors from at least one classification. Searches are performed on all three fields by default.

The Common User Interface

In the following pages of this handout, we will show the main features of the ETRDL system through the Common User and Administrator Interfaces. These interfaces have been designed in order to be highly user friendly. No special knowledge is needed by the user querying the system, the author submitting a new document, or the administrator/librarian responsible for verifying the submission and inserting it into the local collection. On-line helps can be called-up, in English or in the local language, when extra assistance is needed.

The Home Page

The Home Page provides the user with two main options: search/browse any collection or submit/withdraw a document from the local collection.

The search and browse functions can be activated over the entire NCSTRL collection, the ERCIM collection, or the collection of the local institution.

The Home Page is localised by institution, e.g. Figure. 1 shows the Home Page of CNR. The logo of the institution appears in the top left hand corner and a button allows the user to switch between interfaces in English or the local language. The local home pages are implemented according to the particular needs of each partner.

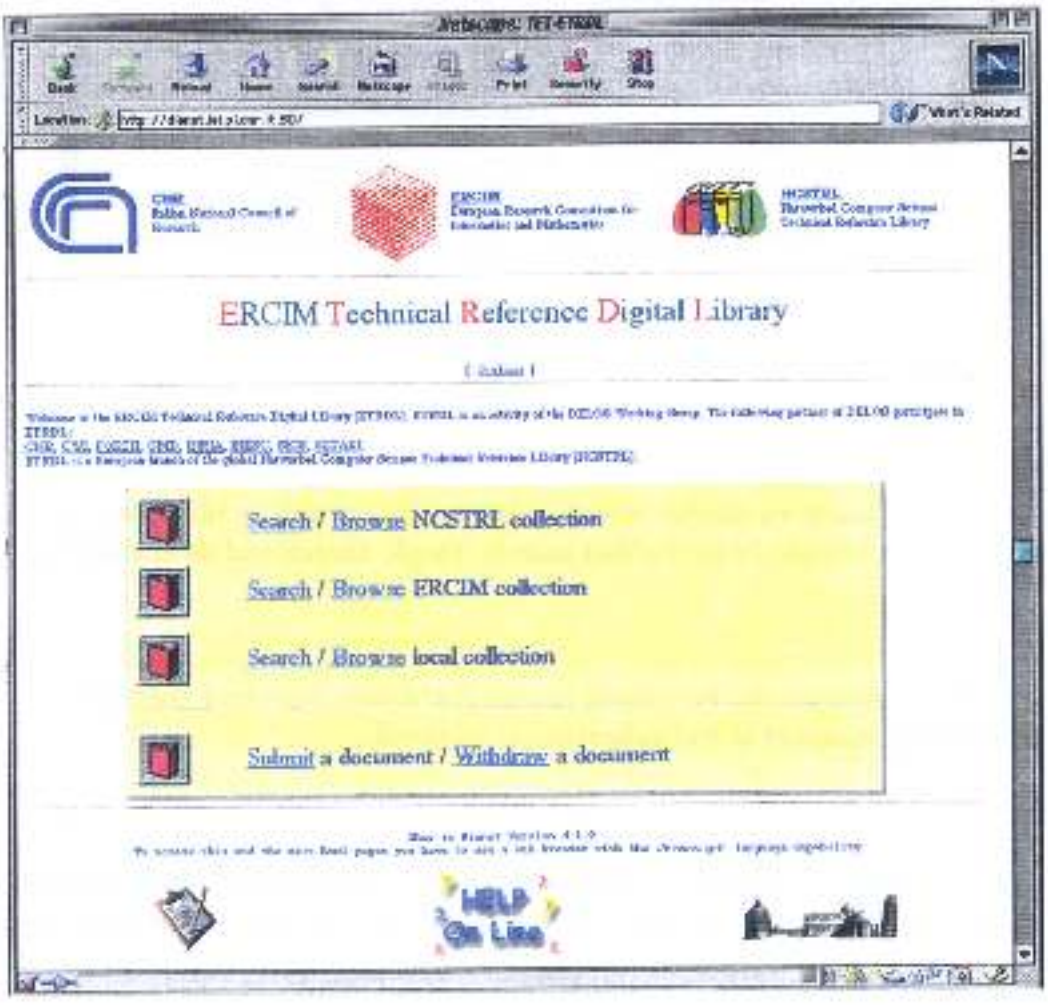

Figure 1

Browsing the ETRDL Collection

The browse function is used to acquire an idea of the content of the collections of the separate ERCIM Institutions. This function in ETRDL has been extended with respect to NCSTRL. Not only can the collections be viewed by year or by author but also by subject classification, (see Figure 2). A document is selected and viewed by clicking on it with the mouse. Documents can be downloaded and then printed.

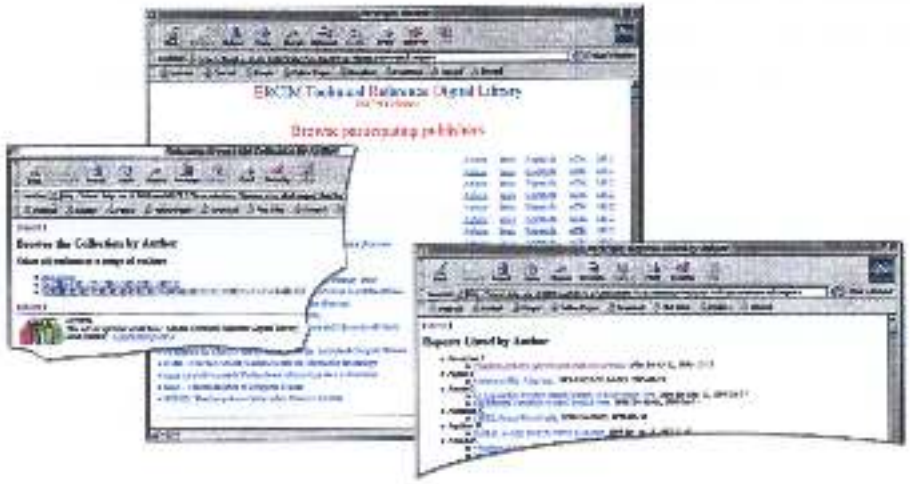

Figure 2

Searching the ETRDL Collection

The user can choose to search either selected collections or all collections. Three kinds of search are available: fielded search, simple search and direct search.

Simple Search

Query terms entered in the simple search field are searched throughout all the fields of the documents in the collection(s) selected.

Direct Search

This field is used when the user want to access a specific document and knows its Document Identification Number, e.g. a librarian wanting to visualise quickly a document.

Fielded Search

The fielded search form for the ETRDL service has four logical components:

1. The bibliographic fields: Title, Author, Subject, Abstract and second language Abstract with a selector to specify the language, and two radio buttons to

specify whether the values entered in the fields should be "ANDed" or "ORed". The selector for the language of the second abstract is only operational if a value is entered in the second abstract field.

2. Three selectors to search documents according to Language, Type and Year.

3. A menu to select one or more collections on which to perform the search, and a check box to select all collections.

4. Two buttons to start the search or to clear the values entered in the fields. The user enters his/her query terms in the field(s) on which the search is to be performed. The value for document type and language are selected from a menu; the default value is All. By default, searches are performed in 'AND' between terms entered in a single field and in 'OR' between fields; the user has the option of changing the default value. An On-line Help is provided to assist the user in formulating the query and specifying search conditions.

Figures 3-6 show screen dumps of the interface for a Fielded Search.

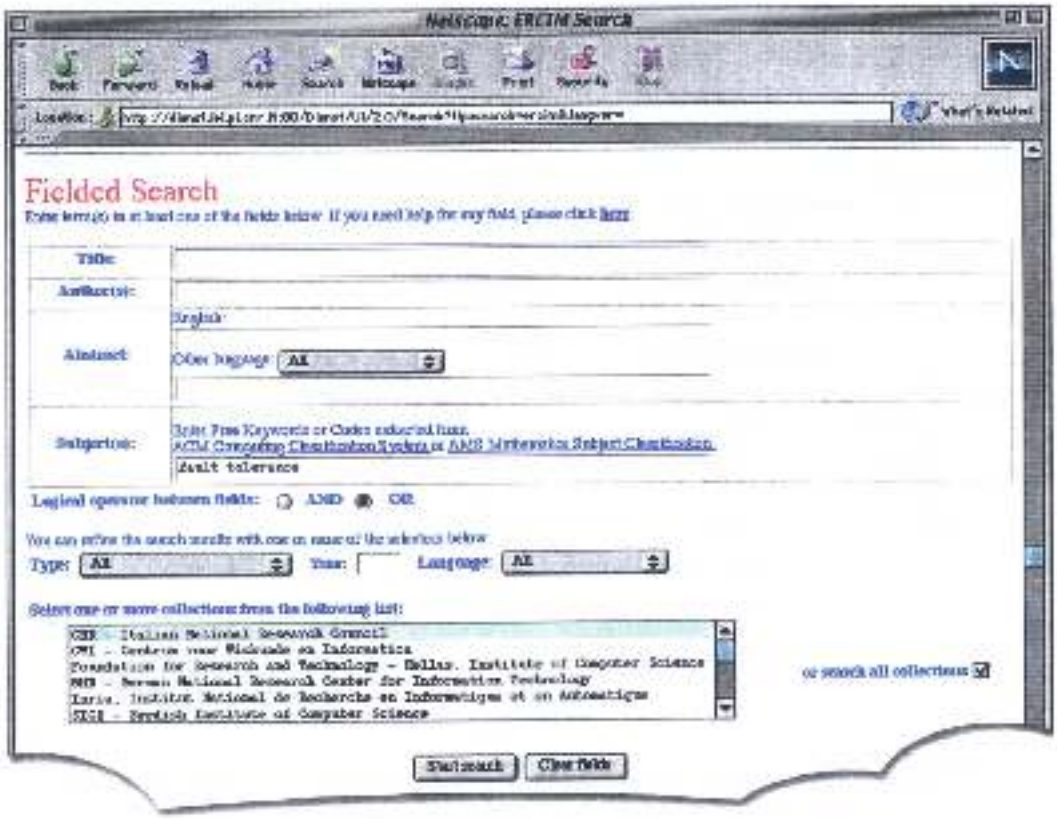

Figure 3

The query entered in the Fielded Search form in Figure 3 is for documents of any type, in any language and from any ERCIM Collection with "fault tolerance" in the subject field. The current local server is highlighted.

Results of the Search

The results of a search are first displayed in summary - the number of documents found is displayed for each publisher.

The documents found are listed, publisher by publisher, the title and authors of each document is displayed. By clicking on a given document the user can view it. Documents may consist of just the title, author(s), abstract and keywords; in this case the entire document is displayed immediately on the screen. If the whole text has been inserted, the user also has the choice of displaying an overview of the document (the whole document in thumbnail format, or page by page). He can also download and/or print out the whole document.

The results for the previous query are shown in Figure 4.

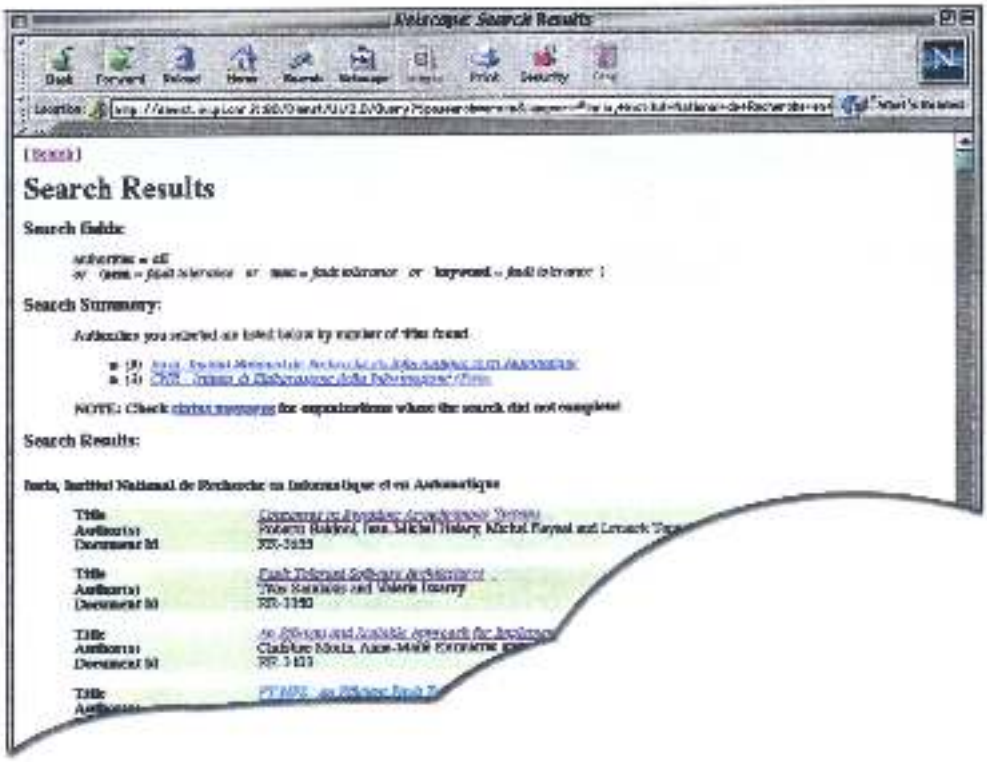

Figure 4

In the next page, Figure 5 shows the bibliographic record and abstract for one of the documents retrieved by the previous query and the instructions for how to view or download the entire document; Figure 6 shows the thumbnail view of the same document.

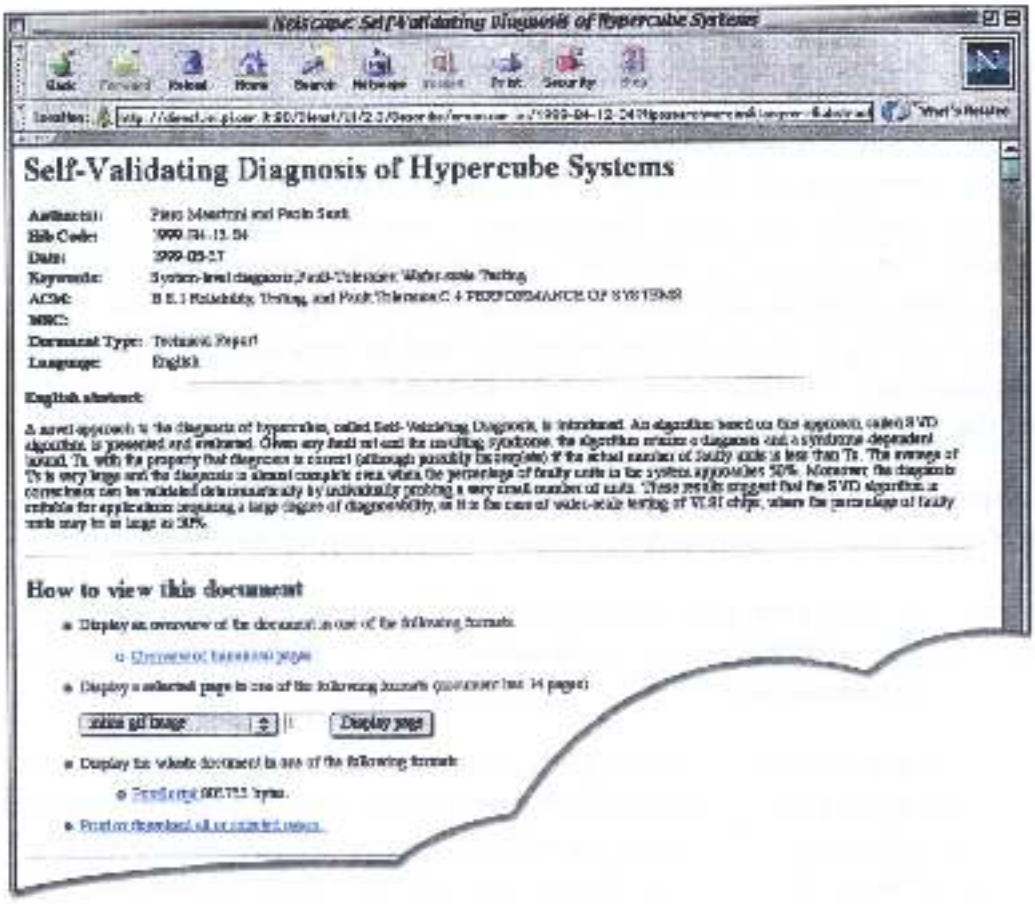

Figure 5

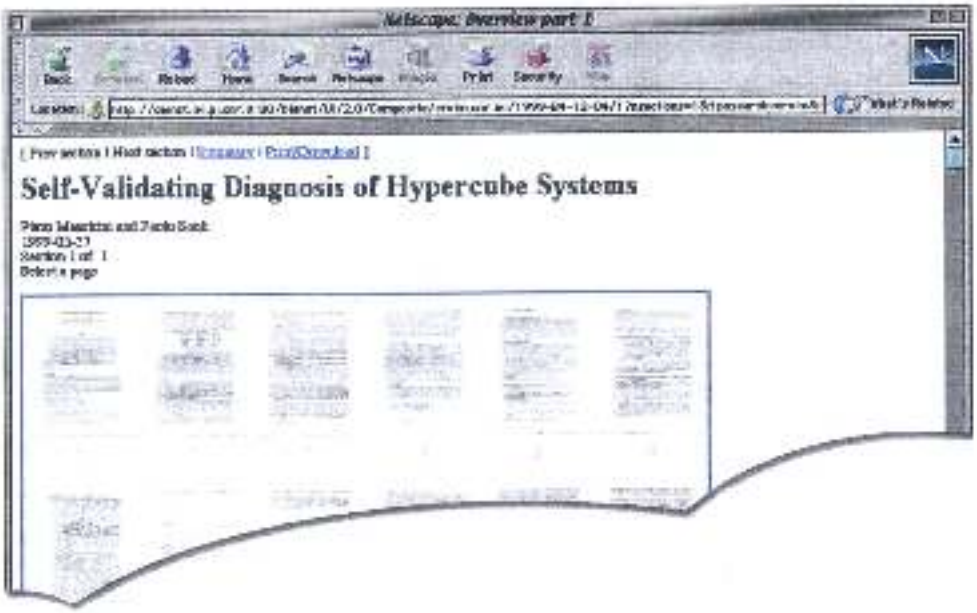

Figure 6

Submitting a Document in the ETRDL Collection

In order to insert a new document in the collection, the document submission form must be completed. All fields on this form are obligatory. For each field, on-line helps are available to assist the compiler. Subjects must be assigned to each document: these should be selected preferably either from the ACM or the AMS classification schemes, which are available on-line. However, the user also has the choice of adding to these or supplementing them by free keywords. If the compiler needs assistance in assigning the correct subjects, he can contact his local librarian using the link at the bottom of the page.

An abstract in English is obligatory for all documents: for documents in a language other than English, an abstract in another language should also be included.

The compiler indicates the name of the file containing the new document to be inserted in the system by using the Browse button to scan his/her system files and mark the appropriate one.

When the form has been completed, the compiler must click the Submit button. The system will display the completed form and request the compiler to confirm that it is correct. If corrections are to be made, the compiler can return to the original form, using the Go Back button. The corrected form is then resubmitted and sent to the system administrator.

On the next page, the submission form can be seen in Figures 7and 8. The compiler has opened a window to the ACM Computing Classification System in order to search for correct classification codes. Figures 9 and 10 show the completed form that has been returned by the system for verification and confirmation.

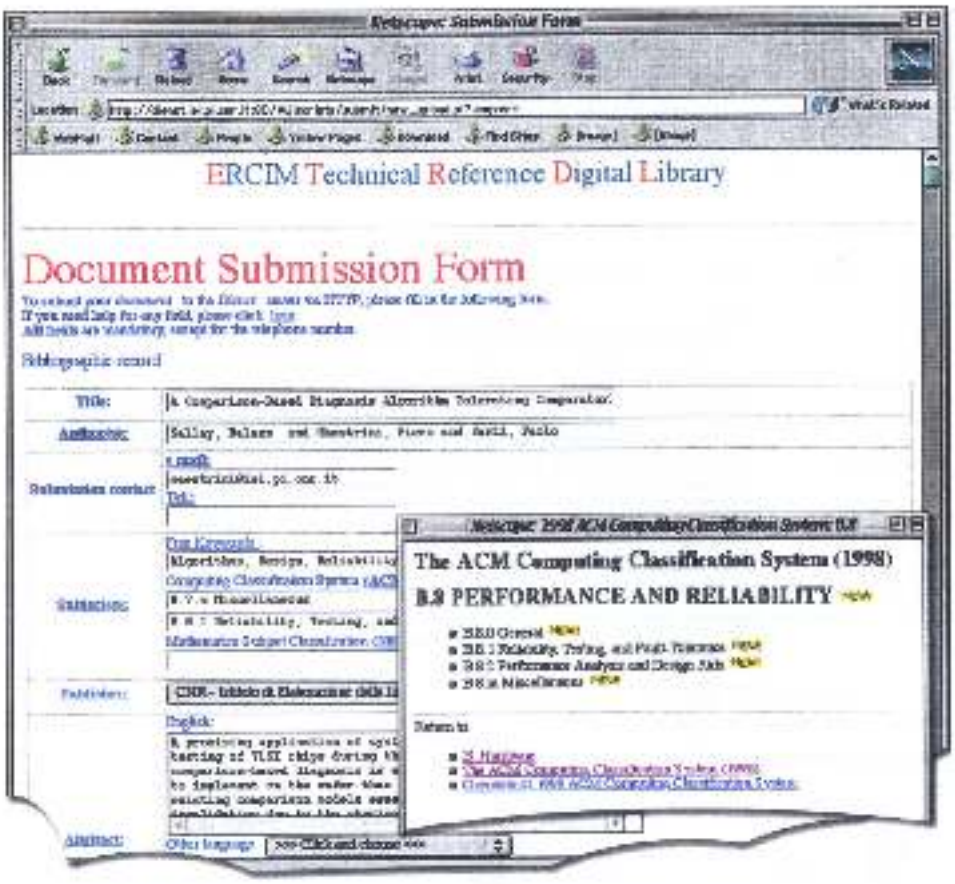

Figure 7

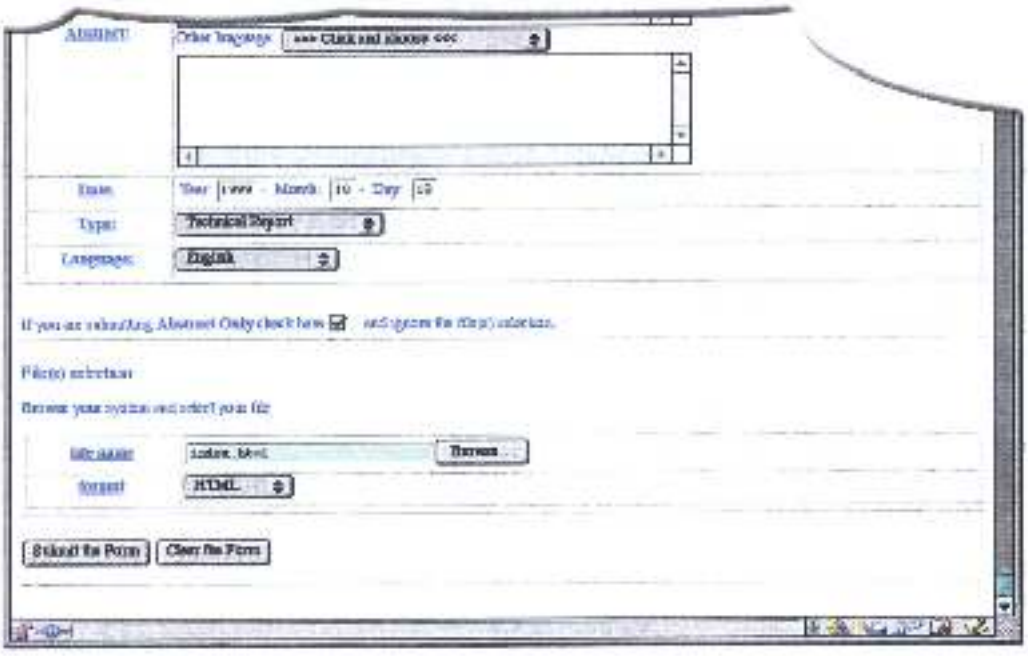

Figure 8

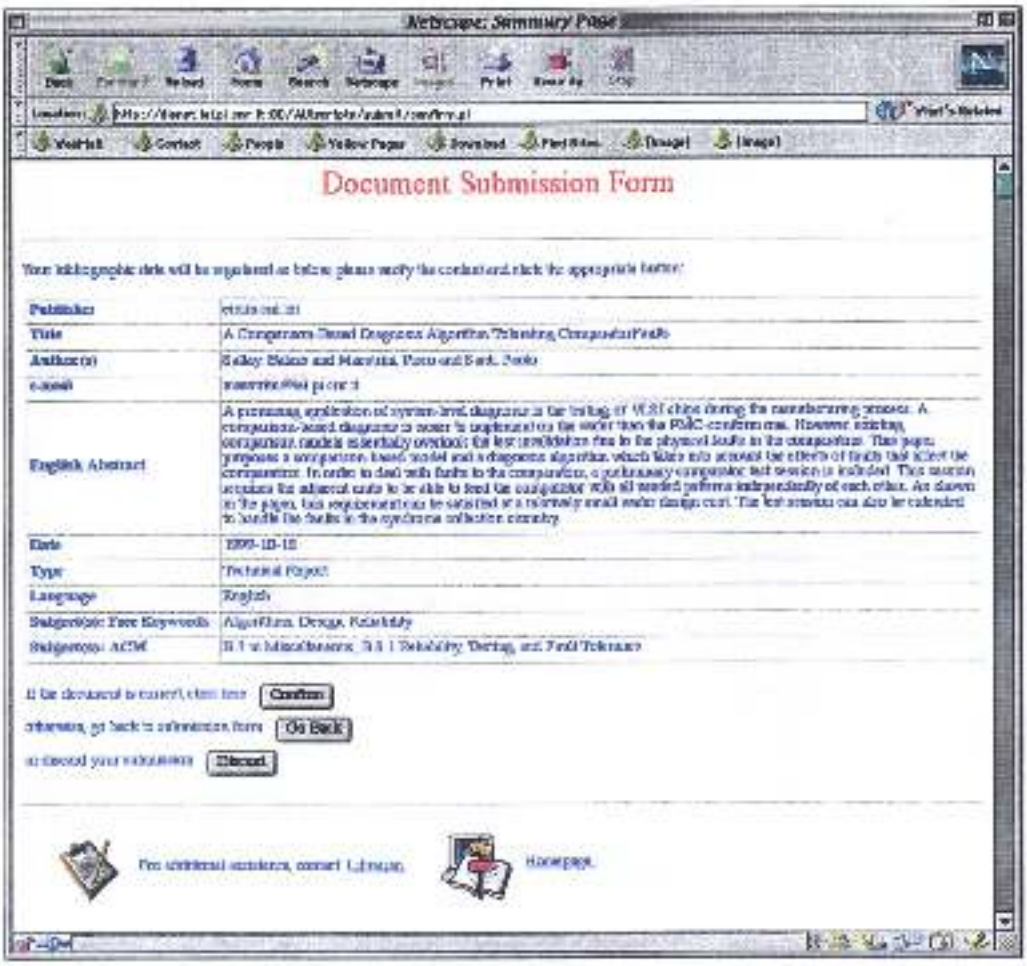

Figure 9

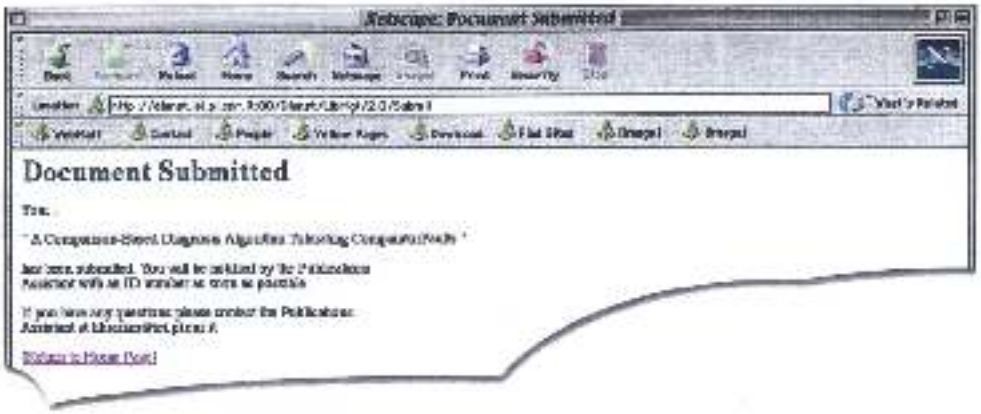

Figure 10

Administration Interface

Each institution uses its own Administration procedures and is responsible for developing an appropriate Interface to be used by the Administrator (often the librarian) for insertion of new or deletion of outdated documents from the collection(s).

The Administrator is generally responsible for assigning the Document Identification Number (DocId.) and verifying the formal correctness of a bibliographic record. Correct records are submitted directly to the system, otherwise the Administrator will contact the compiler asking for the necessary corrections to be made. Figures 11 -13 show screen dumps of the CNR Administrator Interface.

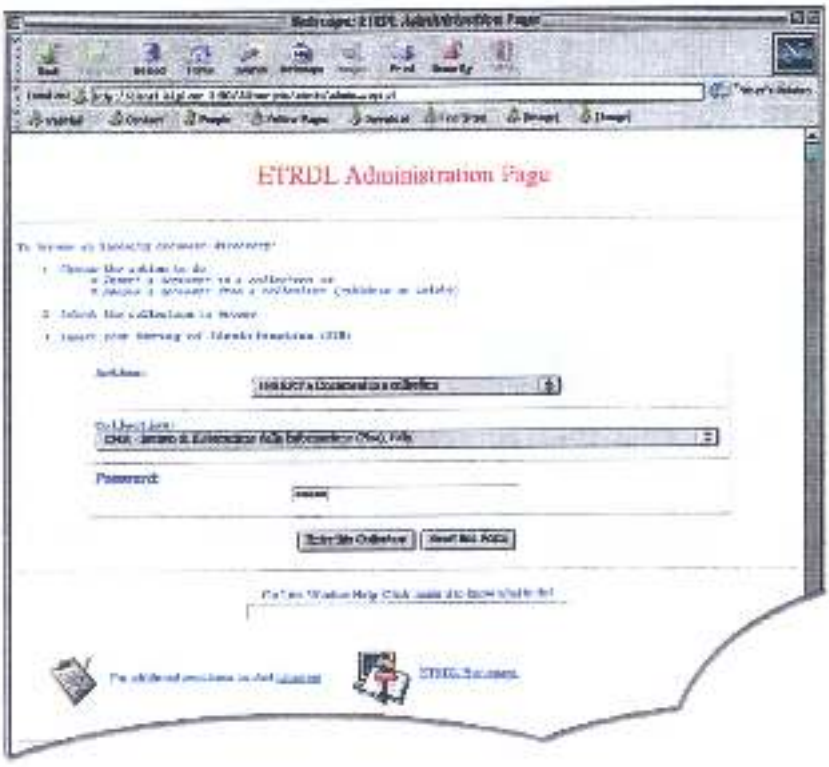

Figure 11

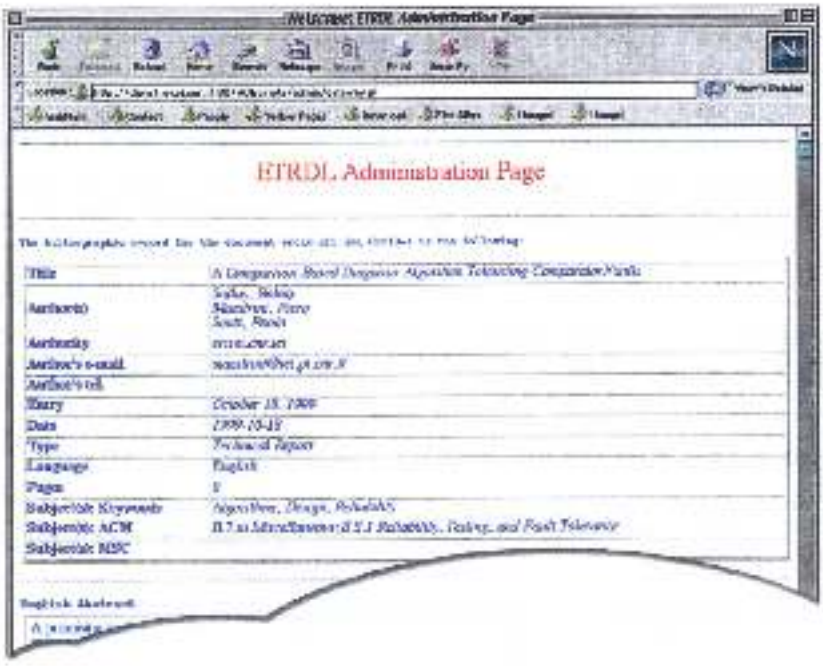

Figure 12

Figure 13

Multilingual Interface

Multilingual Access and Browsing

The version of Dienst used by NCSTRL does not support languages other than English (i.e. no accented characters can be manipulated). The ERCIM Digital Library must be able to cater for the 13 different European languages used by ERCIM institutions. Each national site is responsible for localisation, i.e. implementation of local site user interfaces in the local language as well as English in order to provide user-friendly access for users not familiar with English. Documents included in the system can be in any of the language. The bibliographic record associated with the

document must specify the language of the document and include an abstract in English and an abstract in the language of the document. The user can search for documents in languages other than English by entering terms in the second abstract and selecting the language from the menu (currently Dutch, French, German, Hungarian, Italian, Portuguese, Spanish). Separate indexes are maintained for English and for other languages. The complete Latin-1 character set (ISO_8859-1) is installed so that all diacritic characters can be viewed and searched correctly.

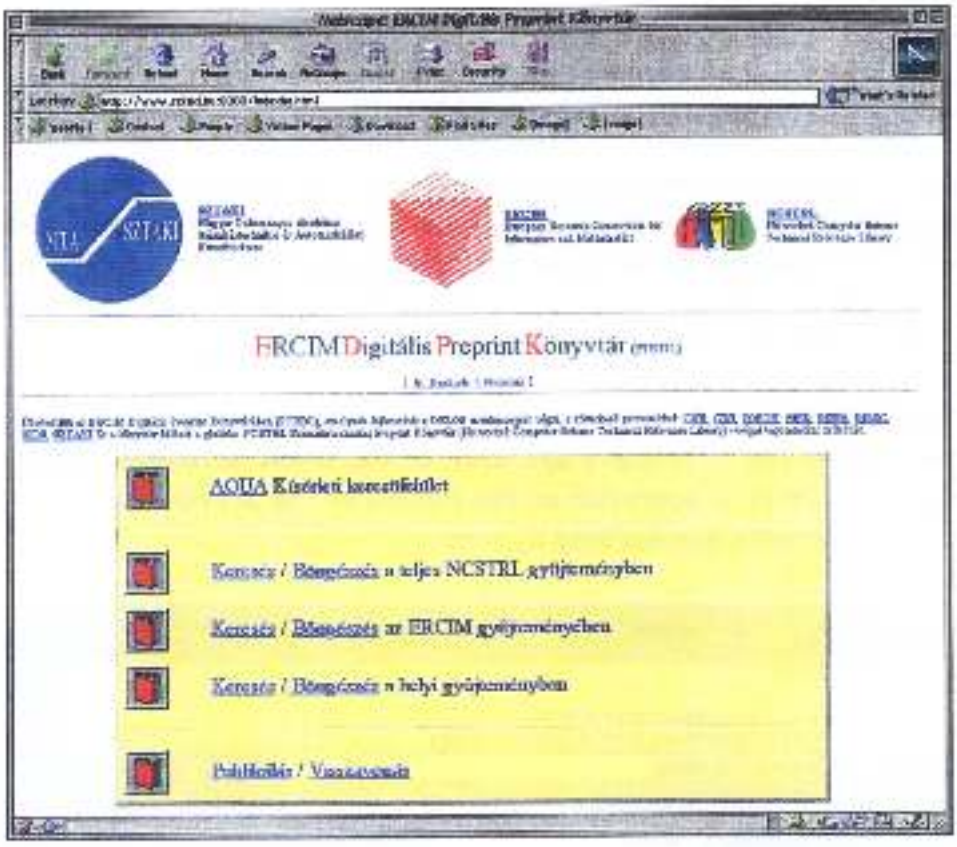

Figure 14

Cross-language Querying

A simple form of cross-language querying is possible using terms from the controlled languages (ACM/AMS). All documents in the ERCIM collection, in whatever language, classified using this scheme, can thus be searched. As all documents must have an abstract in English, English free term searching over documents in any language is also possible.

Using the ETDRL as a Testbed

One of the objectives of ETRDL is to provide project members with a testbed for experimenting new DL tools. Here below we give some examples of past and present research activities at ERCIM institutions aimed at providing internal and external extensions to the basic ETRDL system.

AQUA (Advanced Query User interface Architecture)

MTA SZTAKI, Dept. of Distributed Systems

AQUA is a prototype implementation of a general user interface paradigm which is capable of modelling iterative query refinement and embedding several common and advanced query techniques. The objective of this interface is not only to provide an integrated view for different searching and browsing techniques, but also to help the user to explore the contents of a digital library.

AQUA uses a single list of panels to visualize the user's query refinement steps. The first panel applies the first query operation on the whole document collection, then each panel represents an operation on the results of the previous panel. In this way users can build complex queries step by step.

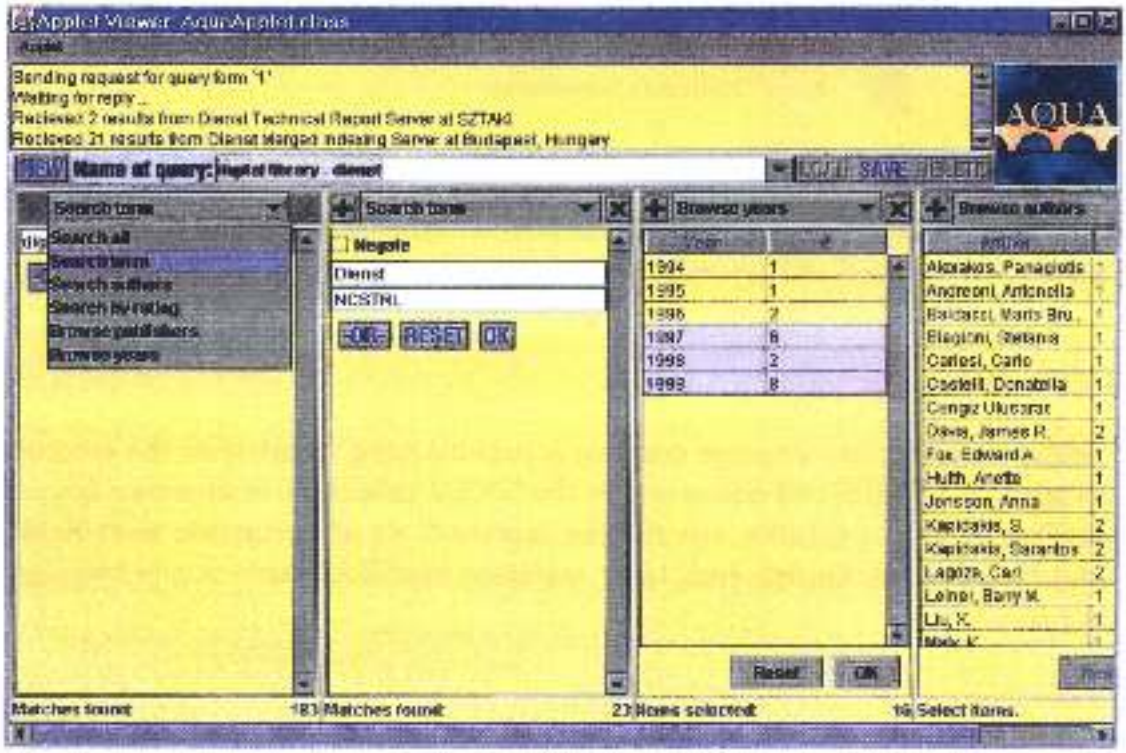

Figure 15

AQUA can be used with the NCSTRL and ETRDL collections, as the two screenshots from an example query demonstrate. The lower part of the AQUA screen contains a scrollable list of panels. In the example the first panel contains a boolean search for the term "digital library", with 183 results from two different digital library servers. The second panel refines the first query by searching for terms "Dienst" or "NCSTRL", and the remaining 23 documents are browsed by publication year in the third panel. The fourth panel shows the authors of documents for the years selected in the third panel.

The next screenshot contains a document view as the final panel. It contains part of the bibliography record and a download menu to view or download the document in the selected format.

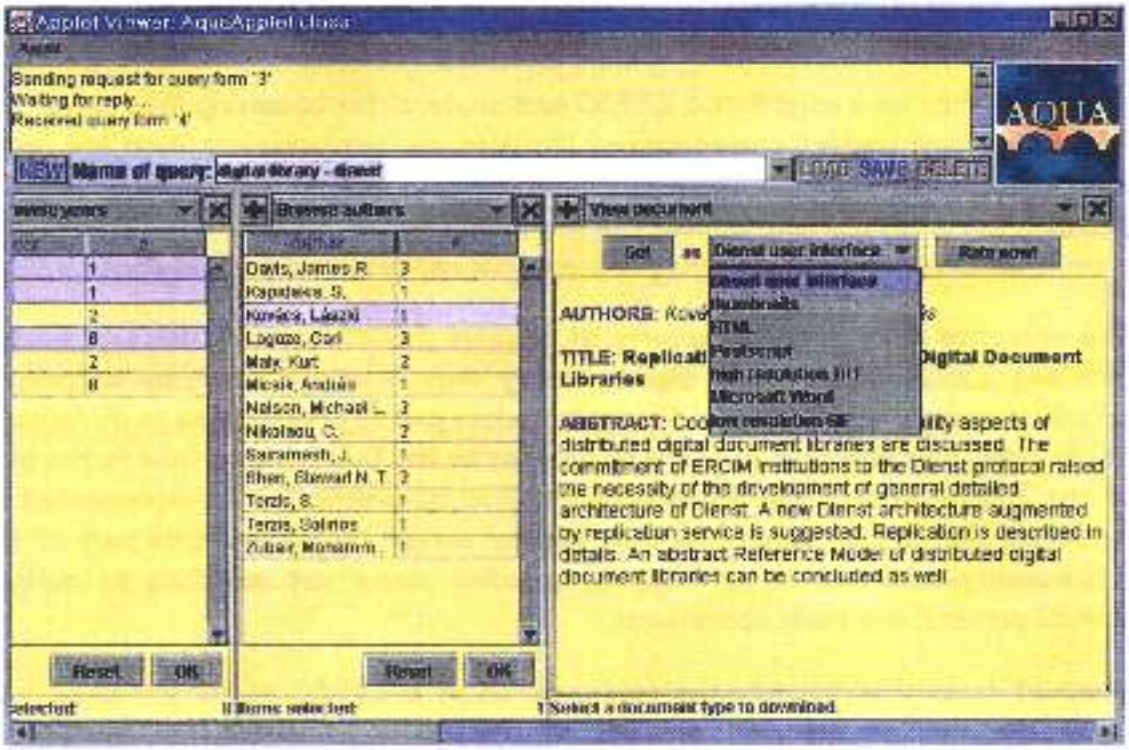

Figure 16

The upper part of the AQUA window serves as a console where the user can follow what is happening within the system and how their inputs are being processed. Between the console and the panel list there are tools to manipulate the whole panel list, queries can be saved and later reapplied to the collection, i.e. each panel's contents are refreshed according to the current information in the digital library system.

Panels in AQUA contain some standard elements: the type chooser is a drop-down select box at the top of the panel, with which the user can change the type of query. The delete button in the upper right corner can be used to remove the panel

from the list. The add button in the upper left corner is for inserting a new panel into the list before the current panel. The status area at the bottom of each panel shows status messages about the progress of the query, about errors in processing and the result. The rest of the panel area is used to visualize the chosen query type. Both the AQUA client and the server has been implemented in Java. The client is an applet which runs in the window of a Java-enabled Web-browser, and shows the actual state of the query, computed by the AQUA server. The AQUA server handles multiple client sessions, and it can access several document collections in parallel. Wrappers are used to connect systems such as Dienst with the server, so different digital library systems can be supported with the same server.

A Z39.50 Gateway

FORTH

Online libraries have established Z39.50 as a protocol for accessing their data. The Z39.50 protocol enables cooperation of libraries, so that users can issue one query and get answers from many libraries. Digital libraries are not just online cataloguing systems, but also hold and provide the material itself online, on a digital form, and provide advanced ways of searching and material retrieval and presentation.

The objective of a research activity at FORTH was to combine the two worlds, providing Z39.50 access to a digital library that is using Dienst, by adapting a Z39.50 server to use the digital library metadata and to provide links to all formats of the data. This approach required no changes to the Dienst server and to the part of the Z39.50 server that implements the Z39.50 protocol; it is implemented by extending the Z39.50 server. An experimental server that implements part of the functionality described in our mapping, including concurrent searching to multiple Z39.50 servers, was made operational.

A set of tools that simplify the configuration of a Z39.50 server attached to a Dienst digital library was also developed.

The LDAP Server INESC

INESC

INESC, an earlier member of the ETRDL project, developed an LDAP service with a multilingual repository for the ACM and AMS classification systems, for integration in the ETRDL system. This multilingual service made cross-language querying possible.

Both users submitting new documents to the collection and those accessing the system to search/browse through the collection can use the LDAP Classification Server.

The Classification Server contains the ACM Computing Classification System and the AMS Mathematics Subject System in English and translations in Portuguese. The directory is designed to host other structured classification systems, and also their translations in multiple languages. The Java applet is completely independent of the contents of the directory, being configured according the information received from it.

When interacting with the Classification Server, the user can select the desired terms, in one or more of the languages, and return to the Collection Server with the terms selected to be used in the task in course. The selected terms can be used in the Collection Server to classify a new document, or to search in the indexes. The actual Classification Server can maintain several indexes, depending on the metadata structure of the collection. One index is supported for each classification schema, but the terms included in these indexes can also be used as generic keywords and indexed in the keyword index (in this sense, a user can perform a free searching in this index using terms from the classification systems).

Additional Search Functionality

CNR

CNR is now studying additional mechanisms for multilingual access and cross-language querying where the user can query in his/her preferred language and retrieve documents matching the query in whatever language they are written.

Project Partners

The ERCIM Technical Reference Digital Library (ETRDL) is one of the activities of the DELOS Working Group, funded in part by the ESPRIT Long Term Research programme (LTR No. 21057), in part by ERCIM, and in part by the partners themselves.

CNR - Consiglio Nazionale delle Ricerche - Italia
CWI - Centruum voor Wiskunde en Informatica - The Netherlands
FORTH - Foundation of Research and Technology - Hellas - Greece
GMD - German National Research Center for Information Technology - Germany
INRIA - Institut National de Recherche en Informatique et en Automatique - France
MTA-SZTAKI - Computer and Automation Research Institute of the Hungarian Academy of Sciences - Hungary
SICS - Swedish Institute of Computer Science - Sweden

Main Contacts

Stefania Biagioni, CNR-IEI (Coordinator)
Frank Roos, CWI
Stella Vourou, ICS-FORTH
Reginald Ferber, GMD-IPSI
Anne-Marie Vercoustre, INRIA
László Kovács, MTA-SZTAKI
Preben Hansen, SICS

For further information on ETRDL, contact Stefania Biagioni (biagioni@iei.pi.cnr.it)

Acknowledgments

ETRDL is a collaborative effort between a number of ERCIM Institutes. Many ERCIM scientists and technicians have been involved in the setting up of the experimental technical reference service.

In particular, we should mention: Jacques Ducloy and Claude Aubrie (INRIA), Sarantos Kapidakis (ICS-FORTH), Antonella Andreoni, Pasquale Pagano, Serena Pisani (IEI-CNR), Lars Nilsander (SICS), Andras Micsik (MTA-SZTAKI).

We must also thank the developers of the Dienst system at Cornell University (USA) and, in particular, Carl Lagoze and David Fielding for their generous assistance and advice.