

8th EAHIL Conference – paper 101

Authors:

Paolo Gardois

Biblioteca Dipartimento di scienze pediatriche e dell'adolescenza, Università di Torino

Gaetano Grillo

Biblioteca Minerva Medica, ASO S. Giovanni Battista, Torino

Chiara Bassi

Biblioteca Ce.V.E.A.S., Modena

Carla Lingua

Biblioteca Dipartimento di scienze pediatriche e dell'adolescenza, Università di Torino

Antonio Piga

Dipartimento di scienze pediatriche e dell'adolescenza, Università di Torino

Monica Daglio

Ce.V.E.A.S., Modena

Planning with users: a catalogue of Internet resources for clinicians, nurses and citizens

Abstract

Background

Many institutions and commercial firms offer lists or catalogues of selected Internet resources. However, end users often complain about resources not properly chosen and described, or catalogues difficult to be handled. It is also vital that catalogues use standard description methods and are scalable, to allow for creation of larger resources by co-ordinating local initiatives. On the contrary many catalogues are not built up on standards and do not take into account user's feedback.

Objectives

A cooperative catalogue of Internet resources for clinicians, nurses and citizens in the pediatric, gynecologic and health promotion field was created. We wanted to verify if user's inputs may enable us to improve the management of the resource.

The analysis of our web server's log file has been used to study the use of the catalogue.

Methods

To build up the catalogue we applied:

- a resource discovery and selection method previously developed from our group (Gardois P, *Evaluating the quality of medical information on the Internet*, 2000 <<http://www.icml.org/posters/post28/poster28.htm>>). A checklist was produced, consisting of 10 main parameters about each resource considered for selection. In addition, brief guidelines were produced describing how to apply these criteria considering local specific needs, mainly in terms of language of the resources, cost for access, adequate balance of resources selected for the different categories of users.

The descriptive method was based on the basic set of 15 Dublin Core elements, with appropriate qualifiers.

These methodologies were tested by a balanced group of final users (20 clinicians, 20 nurses, 20 citizens). These have been selected from lists of junior doctors in Pediatrics and Child and

Adolescent Neuropsychiatry, senior pediatricians and gynecologists, students undergraduating in nursing sciences, senior nurses, citizens.

The tools for feedback evaluation have been questionnaires and data analysis software.

Users tested the usability of the interface, the quality of its descriptions and the resources discovery power.

Results and comments

We aggregated an initial number of about 200 resources. The results presented include quantitative and qualitative evaluation the difficulties in building up the resource, the user judgements and the changes suggested by users. Further developments and open questions include matters related to the quantitative improvement and maintenance of this initiative and the possibility of creating a nation-wide evaluation and description structure about resources in these disciplinary areas.

Summary

Planning with users: a catalogue of Internet resources for clinicians, nurses and citizens.....	1
Abstract.....	1
Background	1
Objectives.....	1
Methods.....	1
Results and comments	2
Summary.....	2
1. Introduction: why another catalogue of Internet resources?	3
2. Selection and cataloguing: problems and methods	4
2.1 Which resources should be catalogued?.....	4
2.2 A checklist for quality evaluation	5
Some sample data about resources retrieved through general directories.....	6
2.3 Resource discovery methods.....	6
2.3.1 Secondary sources used	6
2.3.2 Search methods	6
2.3.2.1 Aims	6
2.3.2.2 Search strategies in single sources.....	7
2.3.2.2.1 Bibliographic databases: Pubmed	7
2.3.2.2.2 Subject browsing in directories	7
2.3.2.2.3 Advanced search on search engines.....	8
2.3.2.3 Saving, managing and reusing search strategies and results.....	8
2.3.2.3.1 Bibliographic databases: data-mining Pubmed results.....	8
2.3.2.3.2 Directories	8
Browsing for subject categories	8
Results management.....	9
The following searches.....	9
Searching archives of newsgroups and mailing lists	9
2.4 The available standards and the Dublin Core choice	9
2.5 Hardware and software.....	12
2.6 Management and financial aspects.....	12
3. Feedback from users.....	13
3.1 Analyzing users' needs	13
3.1.1 systematic revision of the librarians' knowledge of users' needs	13
3.1.2 What users think of our methods: evaluation of the first questionnaire	14
3.1.2.1. Rate of answers to the questionnaires	14
3.1.2.3 Internet resources and search engines recommended by users	16
3.1.3 Users' evaluation of the beta version of ELISIR.....	20
3.1.3.1 Technical issues	25

3.1.3.2 Usability of the user interface	25
3.1.3.3 Users' feedback about planned new functions of the search engine	29
3.3 Continuous post-implementation feedback	31
4. Measuring access and monitoring the system	31
4.1 Usage statistics from Apache log file	32
4.2.1 General statistics	33
4.2.2 Usage of single pages	33
4.2.3 Technical aspects: web browsers and operating systems	34
4.2 PostgreSQL log file	35
5. Conclusions	35
6. Bibliography	36
7. Credits	37

1. Introduction: why another catalogue of Internet resources?

Biomedical professionals and citizens can rely nowadays on a great and growing number of tools in order to retrieve specific biomedical information on the Internet.

These tools can be divided into search engines and directories (often the two are combined in a single resource), which can be either general or specific.

Examples of generic tools could be the health and medicine sections of search engines such as Google or Yahoo or of directories such as the Open Directory Project.

Between the more specific tools, the following free resources are particularly relevant. In table 1 we show the main data about them.

Table 1

Name	URL	Type of catalogued documents and method	Level of description	Number of items on 2002-08-10	Geographical area
BIREME - BVS - Localizador de Informação em Saúde (LIS)	http://lis.bvs.br/xml2html/xmlListT.php?xml%5B%5D=http://lis.bvs.br/lis-Regional/P/define.xml&xsl=http://lis.bvs.br/lis-Regional/home.xsl	Web sites and single documents selected by specialists	Detailed description based on Dublin Core	Not available	South America and English-speaking countries
Cismef	http://www.chu-rouen.fr/cismef/	Web sites and single documents selected by specialists	Based on Dublin Core	10200	French-speaking countries
Cliniweb	http://www.ohsu.edu/clinweb/	Web sites and single documents selected by specialists	Links list sorted by MeSH	About 10.000	Mainly North America
Healthinsite	http://www.healthinsite.gov.au/index.cfm	Web sites and single documents selected by specialists	Resource name, URL, brief abstract, cataloguing date. Based on Dublin Core	8000	Mainly Australia and English-speaking countries
Healthweb	http://healthweb.org/index.cfm	Web sites selected by specialists	Resource name, URL and abstract. No bibliographic formats	Not available	Mainly North America
Karolinska Institutet Library	http://www.mic.ki.se/Diseases/index.html	Web sites selected by specialists	Links with resource name and author/institution	Not available	Mainly Europe and North America
MedHunt	http://www.hon.ch/MedHunt/	Web sites and single documents selected by specialists and robots	Detailed description without indication of bibliographic standards	60.000	Mainly North America and Europe
Medical Matrix*	http://www.medmatrix.org/index.asp	Web sites and single documents selected by specialists	Annotated list: resource name, URL, short abstract, ranking, terms of access	6000 web sites, 1.500.000 documents	Mainly North America
MedWeb	http://170.140.143.133/MedWeb/	Web sites and single documents selected by specialists	List of links divided into sections and sub-sections (no MeSH)	Not available	Mainly North America

MedWebPlus	http://www.medwebplus.com/	Web sites and single documents selected by specialists	Detailed description without abstract or bibliographic standards. Use of MeSH.	26000	Mainly North America
OMNI	http://omni.ac.uk/	Web sites selected by specialists	Description base on ADAM rules adapted for BIOME	About 5000	Mainly English-speaking countries

* Subscription required from 2002-09-01

This table, incomplete as it can be, describes some common characteristics of these tools:

- in most cases, not only web sites are selected, but also single documents, such as guidelines, brochures for patients, etc.
- with few exceptions, these tools are oriented towards the needs of English-speaking users
- all are general biomedical tools, aiming to cover in the same detailed way all the aspects of biomedical sciences
- though all of them use databases to archive resources, only few use standard bibliographic description formats.

On the other hand, considering the specific information needs of users belonging to the institutions participating in the ELISIR project (see paragraph 2.6), they are only partially met by the above mentioned resources. We can summarize these needs as follows:

- our professionals are pediatricians, gynecologists, midwives, therapists and nurses
- citizens need to find information written in Italian, both the content and the search instructions
- our initial hypothesis was that both professionals and citizens would prefer a selection of a limited number of high quality Internet resources with a link pointing to the home page and a good abstract rather than finding a high number of single documents which need a further selection
- finally, the librarians who started the project think that using a standard bibliographic format is a fundamental issue (see paragraph 2.4).

It seems then quite reasonable to create a new catalogue specifically oriented towards our users' needs. This catalogue should be able to link *locally* relevant content with *globally* accepted cataloguing standards

The result was the **ELISIR project: Electronic Library of International Selected Internet Resources**.

At the moment ELISIR is in a beta – test phase and it will be available online from January 2003.

Our project aims to catalogue Internet resources for Health professionals in the fields of Pediatrics, Gynecology, Obstetrics, Nursing and Physiotherapy and for citizens who require specific and reliable information in the biomedical field.

The chosen standard bibliographic standard is Dublin Core (<<http://dublincore.org>). The cataloguing software and the search engine are being developed and they will be released in few months under GPL licence in an international directory such as Sourceforge.

2. Selection and cataloguing: problems and methods

2.1 Which resources should be catalogued?

This point was one of the first to be analyzed at the beginning of the project.

Considering our users' needs, we decided the following policy statement about the types of Internet resources to be selected:

- **exclusion** of e-journals, already available through many catalogues and lists
- **inclusion** of:
 - freely available e-books
 - web sites of scientific and professional associations for health and biomedical professionals
 - web sites of non-profit and volunteers associations of / for patients
 - factual and bibliographic databases
 - subject gateways
 - OPACs of biomedical libraries
 - specific educational web sites, including distance learning and continuing medical education
 - web sites containing documents such as guidelines, images, articles and case reports
 - specific mailing lists, chats, newsgroups and discussion forums
 - web sites of medical faculties and hospitals
 - resources for evidence-based medicine and biomedical research

Resources are catalogued as a whole, without creating different record for their sections, apart from resources hosting various specific resources different from each others. Besides, we have chosen not to catalogue single documents.

2.2 A checklist for quality evaluation

The following step was the definition of a checklist for quality evaluation. This checklist, used by cataloguers, was divided into essential and complementary criteria¹.

Essential criteria state that, in order to include a resource in the catalogue:

- the aim and object of the resource should be clear
- the resource should be relevant for at least one of our groups of users (doctors, other biomedical professionals, citizens)
- the resource should be available on the Internet at no cost, or should be a free section of a greater resource not available for free. Nonetheless, we include resources to which our institutions currently subscribe (apart from e-journals), and also non-free resources who grant a free trial period and have a cost acceptable for a single user
- the language should be one of the following
 - Italian (first choice if the resource is multilingual)
 - English
 - French
 - Spanish
 - German and Arab resources will be considered only if:
 - Provided with at least an abstract in English
 - Present a considerable amount of non-textual data
- sponsors and funding institutions should be clearly indicated
- the resource should guarantee a correct treatment of data collected from users

Complementary criteria should help the cataloguer to decide about the inclusion of a resource in the catalogue, but they alone are not enough to make this decision (i.e., they are less important than essential criteria)

These criteria state that:

- the resource should be created and managed by an important scientific institution
- the person in charge of the resources should be indicated and it should be possible to contact him/her

¹ For an introduction to the problem of selection criteria, see Gardois 2000.

- the resources should be cited in the scientific literature or otherwise well-known and appreciated from specialists
- the pages should be updated and the date of last update should be clearly indicated
- the resource should be compliant with official standards and specifications about the content
- the resource should be usable also with non-updated hardware and software
- graphics and plug-ins should be used in a sensible way

As we show in paragraph 3, we have asked our users an opinion about these criteria. The results are commented in the same paragraph.

Some sample data about resources retrieved through general directories

At the moment, there are no comprehensive data available about the percentage of resources excluded or included using these criteria.

However we can cite two examples.

- Searching the Google Directory (<<http://directory.google.com/> >) on 2002-06-27 for web sites for nurses in Italian, we retrieved 144 records. 20% of them was included in the catalogue. The same search for English web pages resulted in 959 records. From the first 200 analyzed, we included in the catalogue only 16%.
- Searching the same directory for patient education resources in Italian and English, we retrieved an initial number of 100 records. In total 40% were included, but only 20% of the resource in Italian

2.3 Resource discovery methods

2.3.1 Secondary sources used

The following tools were used:

- Bibliographic databases:
 - Pubmed <<http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=PubMed>>
- General search engines:
 - Google, and more specifically Google web directory < <http://directory.google.com/> >
- Specific search engines:
 - Medical matrix < <http://www.medmatrix.org/index.asp> >
 - OMNI <<http://omni.ac.uk/>>
 - BIREME - BVS - Localizador de Informação em Saúde (LIS)<
<http://lis.bvs.br/xml2html/xmlListT.php?xml%5B%5D=http://lis.bvs.br/lis-Regional/l/define.xml&xsl=http://lis.bvs.br/lis-Regional/home.xsl> >

At the end of the beta test phase, the search will include also the following sources:

- CINAHL
- EMBASE
- PSYCINFO
- At least 3 other specific search engines
- Search engines specific for the Italian- French- and Spanish-speaking countries

2.3.2 Search methods

The following methods were tested:

- Search strategies in single sources
- Methods for saving, managing and re-using search strategies and search results

2.3.2.1 Aims

In a first phase of the project, the above mentioned sources were used without a standard methodology, just to assess their resource discovery power.

After that, we decided to follow standard methods, considering the following aspects:

- at the moment, there are no full-time cataloguers in the project
- the maximum amount of time required to thoroughly evaluate a single resource may reach 40'
- the medium time required to catalogue a single resource is about 20'
- analyzing and selecting resources from the above mentioned sources is a very time-expensive activity

Thus, we outlined the following aims:

- avoiding to waste time on resources which won't be catalogued
- consequently, starting the selection process from set of records retrieved from very reliable sources
- sparing the time required by activities which can be automated, such as repeated typing of search strings in engines, browsing from the beginning sections of directories several times, etc.

2.3.2.2 Search strategies in single sources

2.3.2.2.1 Bibliographic databases: Pubmed

When bibliographic databases are used, the first thing we do is a conceptual analysis of the object of the research, which allows to correctly define the problem and consequently the search string(s).

Since we need to retrieve records about articles describing Internet resources, we face two different possibilities:

- the bibliographic data cite the address of the resource
- the bibliographic data *do not* cite the address of the resource

For each one of the two possibilities we outlined a basic search string, which may possibly be combined with more specific search terms:

a)

www[Text Word] OR http OR mailto OR telnet OR ftp[Text Word]

Allows to retrieve records containing terms related to the address of Internet resources

b)

("search engine" OR "search engines" OR "data bank" OR "data banks" OR database OR databases OR "subject gateway" OR "subject gateways" OR clearinghouse OR repository OR directory OR portal OR forum OR "mailing list" OR "mailing lists" OR chat OR listserv OR irc)

AND

(internet OR "web site" OR "web sites" OR "web resource" OR "web resources" OR "world wide web")

NOT

(www[Text Word] OR http OR mailto OR telnet OR ftp[Text Word])

Allows to retrieve records about Internet resources which do not cite their address, without retrieving again the records already discovered using the string a).

The search results of both strings have to be analyzed manually for the first time.

2.3.2.2.2 Subject browsing in directories

Searching bibliographic databases, though useful to retrieve relevant resources, is not enough. Indeed, it is necessary to use also general and specific directories and search engines to retrieve resources not cited in journal articles or available in languages other than English.

Searching these sources, firstly we have to define the specific sections to be used, in order to perform the same search as we come back to a single source to update the catalogue after a period of time.

After this initial task, we browse through the directory selecting the most important resources.

2.3.2.2.3 Advanced search on search engines

After some testing, we concluded to limit our search to the directory, and not to query directly the whole search engine.

This strategy restricts the number of resources which could be retrieved, but limits the time required to analyze the results, because web sites cited in the directory were preselected by experts.

In the future, we'll reconsider the problem, in order to decide whether to use the advanced search in a complementary way, to retrieve resources in French, Spanish or Italian search engines.

2.3.2.3 Saving, managing and reusing search strategies and results

2.3.2.3.1 Bibliographic databases: data-mining Pubmed results

Search results from strings a) and b) (see paragraph 2.3.2.1) are saved as text files.

From the record set obtained with a) we extract the lines containing the addresses, using *grep* software for Unix/Linux.

Afterward, we refine the lines obtained using the *gawk* software for Unix/Linux, extracting only the addresses, which are sorted alphabetically.

Subsequently, we launch automatically the single HTTP addresses using the HTTP method HEAD, and all addresses giving as a result a 404 error (page not found) are discarded.

The addresses contained in this final list are viewed one by one by cataloguers.

Results obtained using string b), instead, are directly viewed one by one and not saved. Internet resources retrieved will be analyzed and then included or discarded and put into the list of discarded resources.

After a given period of time, both strings (saved as URLs) are relaunched on the bibliographic database, selecting only the records loaded in the database after the date of the previous research.

If the search results are not too many, they can be viewed directly. Otherwise, it could be useful to extract the URL with methods previously described and to compare the new list with the archived list (see following paragraph).

2.3.2.3.2 Directories

Browsing for subject categories

Firstly, the URLs of all the subdirectories in which resources were searched are saved in a list (text file), which becomes the basis for the next research.

Results management

Web pages contained in a single subdirectory of the main directory are processed with *grep* and *gawk* to obtain a list of URLs. The list is saved in a text file.

The following searches

After a given period of time, all the relevant subdirectories are browsed again and their content is saved in a file text, which is processed using the same method as before (*grep* and *gawk*) to obtain a list of URLs.

Finally, using *diff* for Unix/Linux the differences between the old and the new list of URLs are printed, generating a new text file containing only the new addresses loaded into the directory's database in the time period between the first and the second search.

Obviously, these two tasks will be fully automated.

It will thus be possible to view only the new address, sparing a lot of time.

The only other thing to do manually at every catalogue update will remain a check of the subdirectories of the main directory, to identify new relevant subdirectories.

Searching archives of newsgroups and mailing lists

"Internet" and "the web" are not synonyms. Web sites are by far the most important Internet resources, but other relevant resources are available

While FTP and TELNET resources are generally retrievable through search engines, this is not the case for mailing lists and newsgroups.

Since these two types of resources could be precious interactive tools both for professionals and for citizens, mainly to give and receive advice and suggestions about particular problems, we decided to include them in the ELISIR project.

Newsgroups are searched through the following search engines:

- Mailgate.ORG Web Server < <http://www.mailgate.org/> >
- Google Groups (was: Dejanews) < <http://www.google.com/grphp?hl=en&ie=ISO-8859-1&q=> >

To retrieve mailing lists, it is useful the following engine:

- CataList, the official catalog of LISTSERV® lists < <http://www.lsoft.com/catalist.html> >

2.4 The available standards and the Dublin Core choice

From a librarian's point of view, an Internet resource is a document (or a set of documents) and a catalogue is a way (perhaps *the way*) to make resources available to users: thus, it seems necessary to apply bibliographic control to this particular kind of documents.

Besides, using a cataloguing standard widely accepted by the international community makes it easier to integrate in the future bibliographic records from many single catalogues, built on the same standard.

Finally, using an easily convertible standard will ease the task of integrating records created with different standards.

As far as Internet resources are concerned, the choice of a particular standard for bibliographic description, for author and subject headings and for classification is difficult, due to the number of different standards available.

Summarizing, an Italian project aiming to catalogue biomedical Internet resources could choose between the two following alternatives:

- Using ISBD(ER) for bibliographic description, RICA (Italian rules for author headings) for author headings, MeSH as a thesaurus and DDC for classification, possibly integrating all these standards in a single bibliographic format such as UNIMARC
- Using Dublin Core (DC) qualified by MeSH and possibly DDC and local elements, using the RICA for choosing the correct author headings and their right form.

The second alternative was chosen, after considering carefully the following issues:

- Compared to ISBD(ER), DC is a more concise standard, but shows a good modularity and flexibility. Various available schemes can be added to unqualified DC.
- DC is the standard chosen by many international projects and is the architectural framework for the Open Archives Initiative (OAI)²
- For our project, mainly built on volunteers' work, DC seemed to be the better compromise between description accuracy and an acceptable speed, also keeping in mind the volatility of Internet resources
- Finally, DC is convertible to UNIMARC and – through it – to ISBD. Besides, it can be well adapted to a FRBR oriented catalogue³

In our project we have adopted the base structure of DC and some appropriate qualifiers. In the cataloguing software, all the elements outlined below were implemented though some (e.g. DDC and SOURCE) are not used at the moment.

The following table summarizes how we adapted DC to the needs of our users.

Table 2

Element	Qualifiers	Element Encoding Scheme(s)
Title	Alternative title	
Creator		
Subject	MeSH ; DDC (not used)	MeSH (at present, in English; MeSH in Italian will be added as soon as they are published) ; DDC
Description	Table of contents (used only when a summary of the resource is available in the resource itself); Abstract	
Publisher		
Contributor		
Date	Created ; Valid ; Available ; Issued ; Modified	W3C-DTF (YYYY-MM-DD)
Type		DCMI Type Vocabulary
Format	Medium	IMT
Identifier		URI
Source (not used)		URI
Language		RFC 1766
Relation	Is Version Of ; Has Version ; Is Replaced By ; Replaces ; Is Required By ; Requires ; Is Part Of ; Has Part ; Is Referenced By ; References ; Is Format Of ; Has Format	URI
Coverage	Spatial Temporal	TGN W3C-DTF (YYYY-MM-DD)
Rights		

As already stated, RICA were used for choice and form of author headings, corresponding to CREATOR and CONTRIBUTOR.

Beside, for meeting local needs, the following elements were added

² Though created by the e-prints community, the OAI aims to involve in the projects a wide range of documents and meta-documents, and among them also bibliographic records. Besides, cataloguing different resources using DC metadata (inside or outside the document itself) allows for a good integration of bibliographic data and full text documents inside a single portal. An example could be a web site which might contain the present catalogue and also papers, case reports, biomedical images, etc.: a uniform DC metadata structure (even if DC could be differently qualified for different type of resources) would allow for a unified search in the different archives using, say, a MeSH term for a single disease.

³ See Fasella 2001.

- **Users (Utenti):** specifies the type of users for which the resource is advised (doctors, other health professionals of citizens)
- **Notes:** used by cataloguers to make short temporary notes about the resource catalogued
- **Restricted access (Restrizioni d'accesso):** used if the access to the resource is restricted only to registered users or to some particular category of users
- **Cataloguing date (Data di catalogazione):** date in which the resource was catalogued or the record was last modified
- **Cataloguer (Catalogatore):** name of the cataloguer

Here's an example of a resource catalogued using our qualified DC:

Table 3

TITLE	American Academy of Nurse Practitioners
URI	http://www.aanp.org/
ALTERNATIVE TITLE	AANP
CREATOR	American Academy of Nurse Practitioners : AANP
SUBJECT	Nursing
DESCRIPTION	<p><i>Table of contents:</i> Home ; Table of Contents ; About the Academy Certification Program ; Conferences and Meetings ; CE Calendar ; Directory ; Employment ; Fellows ; Foundation ; Foundation Store ; Grants ; Hotline ; International Journal ; Legislative Information ; Links ; Membership ; Mission Statement ; NP Program Directory ; Position Statements ; Practice Information ; Publications ; Public Relations ; Information ; Regulatory Information ; Recruitment ; Scholarships ; State Awards ; White Pages.</p> <p><i>Abstract:</i> Sito dell'American Academy of Nurse Practitioners. Può servire agli infermieri professionali di tutte le specialità ma anche ai loro pazienti e ad altri operatori della salute. Contiene "Hot Links" con un'ampia schiera di risorse relative all'informazione sulla pratica clinica e sulla legislazione. La sua pubblicazione è accessibile solo ai membri dell'associazione.</p>
PUBLISHER	
CONTRIBUTOR	
DATE	Date created: Date issued: Date modified: 2002-07-09 Date valid (start): Date valid (end): Date available (start): Date available (end):
TYPE	Text.Homepage.Organizational
FORMAT	text/html
SOURCE	
LANGUAGE	en
RELATION	Is Version Of: Has Version: Is Replaced By: Replaces: Is Required By: Requires:

	Is Part Of: Has Part: Is Referenced By: References: Is Format Of: Has Format:
COVERAGE	United States
RIGHTS	© American Academy of Nurse Practitioners. All rights reserved.
UTENTI	infermieri cittadini
NOTES	E' consigliato per una migliore visione del sito Microsoft Internet Explorer 5.5
RESTRIZIONI D'ACCESSO	Il FT di The Journal of the American Academy of Nurse Practitioners(JAANP) è accessibile solo per i soci come benefit.
DATA DI CATALOGAZIONE	2002-10-07
CATALOGATORE	Grillo

2.5 Hardware and software

ELISIR works on a IBM server xSeries 230 with Pentium 3 1000MHz, 256 Mbyte RAM, 3 hard disk UltraSCSI.

About the software, the ELISIR has adopted from the beginning the open source philosophy⁴. Whenever possible, with some small exception, the project uses open source software, and all the software produced will be released under a GPL licence⁵. Thus, our cataloguing software will take advantage of international collaboration for improvement and amendments.

The server runs Linux Mandrake 8.0 as operating system. The web server is Apache-AdvancedExtranetServer 1.3.20. The DBMS used for cataloguing is PostgreSQL 7.1.2 and the web interface to the database was built in PHP.

2.6 Management and financial aspects

The ELISIR project is coordinated by the library of the Pediatric and Adolescent Medicine Department of the University of Turin, Italy (Biblioteca del Dipartimento di Scienze Pediatriche e dell'Adolescenza dell'Università di Torino, < <http://www.pediatria.unito.it/biblio/> >). The partners are Minerva Medica library of the S. Giovanni Battista Hospital in Turin (Biblioteca Minerva Medica dell'ASO S. Giovanni Battista di Torino) and CeVEAS library in Modena <<http://www.ceveas.it/ceveas/viewPage.do?idp=2>>.

ELISIR is a two-year project, part of the wider project *The web for scientific information for pediatricians and gynecologists* ("Il web per l'informazione scientifica in campo pediatrico e ginecologico") co-sponsored by Compagnia di S. Paolo, Turin < <http://www.compagnia.torino.it/> > and Pediatric and Adolescent Medicine Department of the University of Turin (Dipartimento di Scienze Pediatriche e dell'Adolescenza dell'Università di Torino < <http://www.pediatria.unito.it/biblio/> >).

4 cataloguers belonging to the 3 institutions take part in the project; no one of them working full time on it. An IT expert is in charge of software development.

⁴ For further information, see *GNU's Not Unix!* <<http://www.gnu.org/>> and *oss4lib - Open Source Systems for Libraries* <<http://www.oss4lib.org/>>. From the ever-growing literature on the subject, we only select Eric S. Raymond, *The Cathedral and the Bazaar* <<http://www.tuxedo.org/~esr/writings/cathedral-bazaar/>>.

⁵ See: GNU General Public License <<http://www.gnu.org/copyleft/gpl.html>>.

3. Feedback from users

When planning to develop a web resource, it is vital to take into account users' feedback from the very beginning.

Literature⁶ particularly recommends to consider users' information needs, in order to be able to plan contents and structure of the site in an efficient way.

In our project, we began by studying the following model:

- Analysis of users' information needs
- Translation of these needs in tasks to be performed using the system
- Rapid prototyping of an structure and user interface which could allow to perform these tasks
- Evaluation of the prototype by users
- Identifying and finding solutions to problems starting from users' feedback
- Final implementation, subject to periodical revisions

ELISIR has adapted these general scheme to local needs in the following ways:

- In the planning and beta-test phases of the development of the cataloguing software and the search engine:
 - analysis of users' information needs by questionnaires and systematic revision of the librarians' knowledge of users' needs, acquired through their activities as reference librarians and teachers
 - creation of a resource selection method, a cataloguing system and a search engine based on the assessed users' needs
 - users' evaluation of these three issues (method, cataloguing system and search engine)
 - Amendments to the systems using beta-testers suggestions
- When the system will be fully implemented:
 - continuous users' feedback

Aim of this paragraph is to share some of the useful experience acquired analyzing users needs and using beta-testers' feedback to amend the system.

3.1 Analyzing users' needs

3.1.1 systematic revision of the librarians' knowledge of users' needs

Discussing and systematising their previous knowledge, librarians involved in the project have identified the following users' needs:

- using Internet resources different from the traditional ones (catalogues, bibliographic databases, e-journals). A growing number of users requires guidance to discover Internet resources such as guidelines clearinghouses, image or case report databases, web sites of professional associations, etc.
- we assumed that users lack a unified resource usable to retrieve such information. Thus, users are forced to waste their time retrieving resources of interest
- users consider the Internet also as *forum* and *agorà*. The net is not only the greatest (virtual) *library* in the world, but also a place where professional and citizens can meet and share opinions. However, it is often difficult to find the right virtual place for specific biomedical arguments.
- Both professionals and (especially) citizens require resources previously selected by specialists. This allows firstly to spare time and then also avoids problems who could occur to people who are not expert enough to evaluate the quality of a resource. It is clear, however, that users should be educated to accurately analyze and criticize also previously selected resources.

⁶ See Kinzie MB et al. 2002

3.1.2 What users think of our methods: evaluation of the first questionnaire

The four above mentioned hypotheses were verified by means of a first questionnaire, to which users answered in May 2002, before librarians set up the first beta version of the cataloguing software and tried to catalogue the first group of resources.

For both questionnaires we have built a group of beta-testers composed of

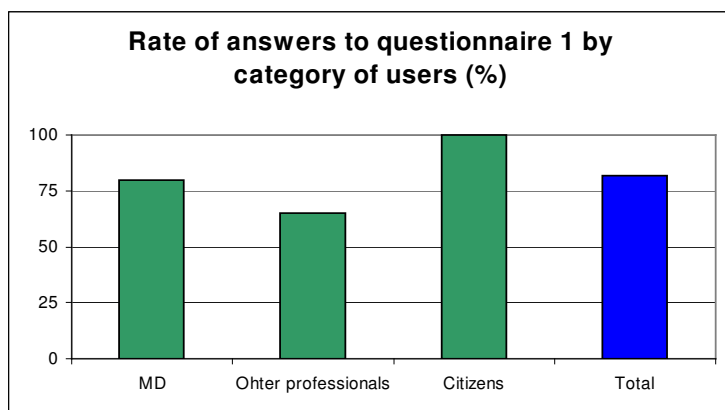
- 20 medical doctors (9 senior pediatricians and gynecologists and 11 junior doctors) working at OIRM S. Anna hospital in Turin
- 20 non-medical health professionals (11 nurses, 6 physical therapists, 2 experts in genetics, 1 midwife) working at OIRM S. Anna and S. Giovanni Battista hospitals in Turin
- 20 citizens selected from the participants to the CeVEAS project “Laboratory for the competent citizen” (“Laboratorio Formativo per il Cittadino Competente”) in Modena.

The only necessary requirement to be included in these groups was a previous, also very small experience of browsing the Internet. The aim was to select users having at least a little idea of what kind of problem was to find information on the Internet.

The next paragraphs show the main data acquired through this first questionnaire.

3.1.2.1. Rate of answers to the questionnaires

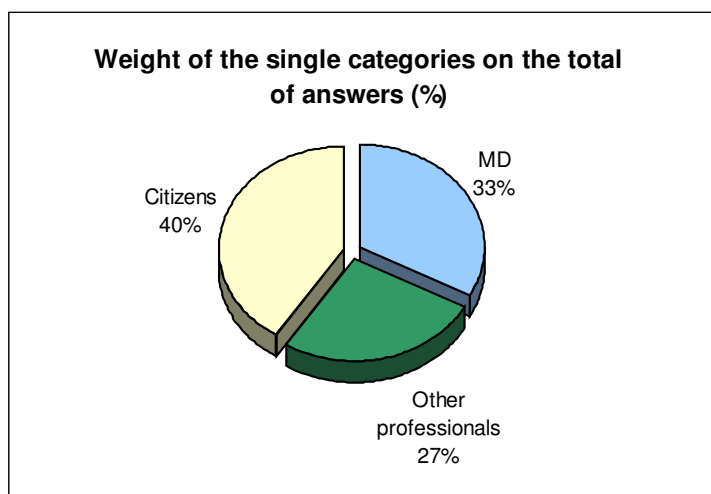
Figure 1



The highest answer rate by category has been reached by citizens.

Figures 1 and 2 must be considered when interpreting the data shown below, especially because the “weight” of the different categories in the total of answers is not exactly the same (see figure 2).

Figure 2



3.1.2.2 Selection criteria evaluated by users

Data from this questionnaire show that users and librarian substantially agree about the criteria to be used to evaluate Internet resources.

No user has suggested new criteria, while most of the answers (70-80% for essential criteria and 60-90% for the complementary ones) consider “very important” the criteria we proposed. Only a very small part of the answers (less than 10%) considers these criteria “not important at all”.

The “most criticized” essential criteria were “sponsors and funding institutions should be clearly indicated ” (20% of answers were “not important at all”) and the criteria about the language of the resource (8% “not important at all”).

The first issue indicates that we should try to do our best to help especially non medical professionals and citizens to become more aware of the problem of “hidden” sponsors and conflict of interests in biomedical literature and information, while medical doctors seem more aware of the problem.

We decided, therefore, to create a file to be linked to the search engine which explains why we have chosen these particular criteria, with further details about this one.

The second issue requires careful consideration about the opportunity to include also resources in German or Arab – the two languages users proposed to add to the criteria.

We concluded to include in the catalogue only German and Arab resources with at least an abstract in English and a significant amount of non-textual data (images, sound, etc.).

About the complementary criteria, users 12% of users answered “not important at all” only about “graphics and plug-ins should be used in a sensible way”: other criteria about IT aspects of resources, though, were considered “very important”.

Figure 3

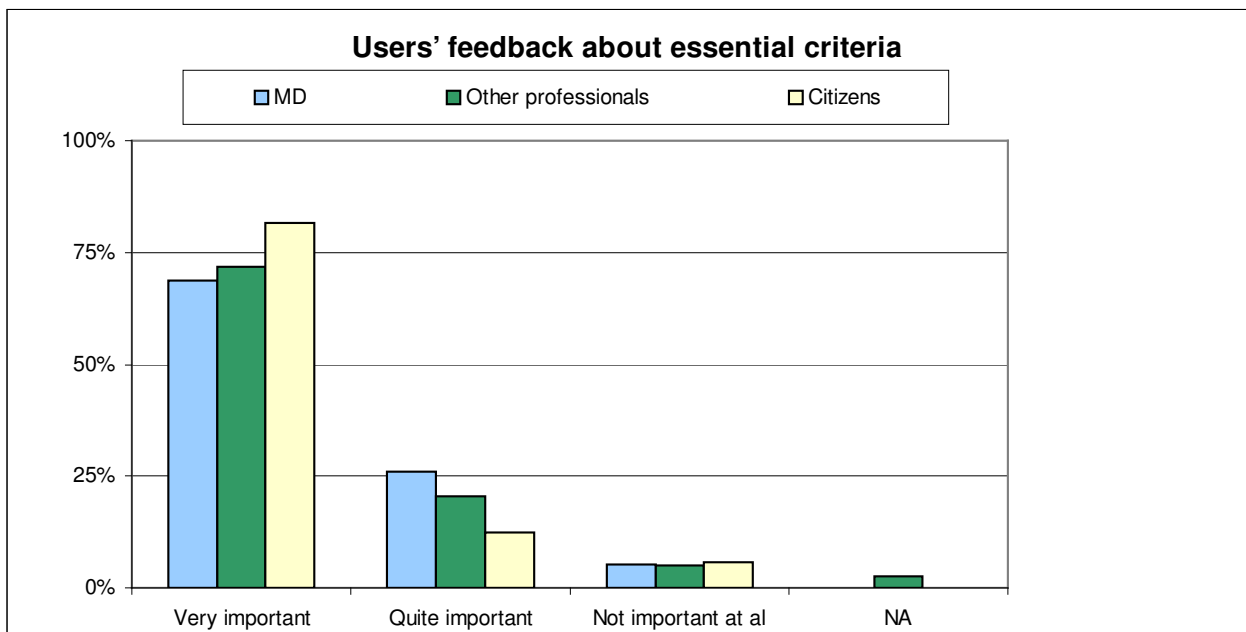


Figure 4

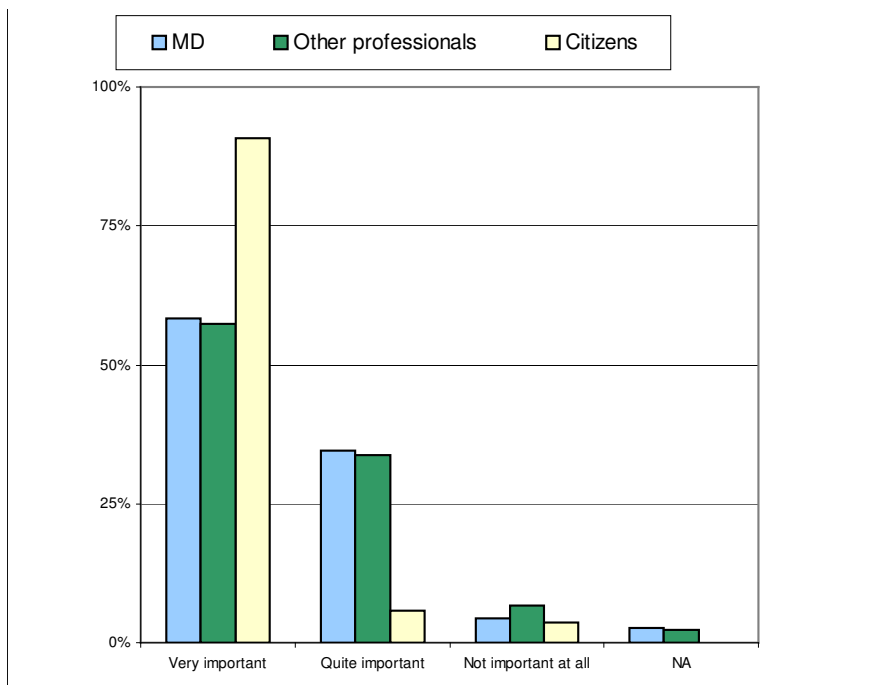


Figure 5

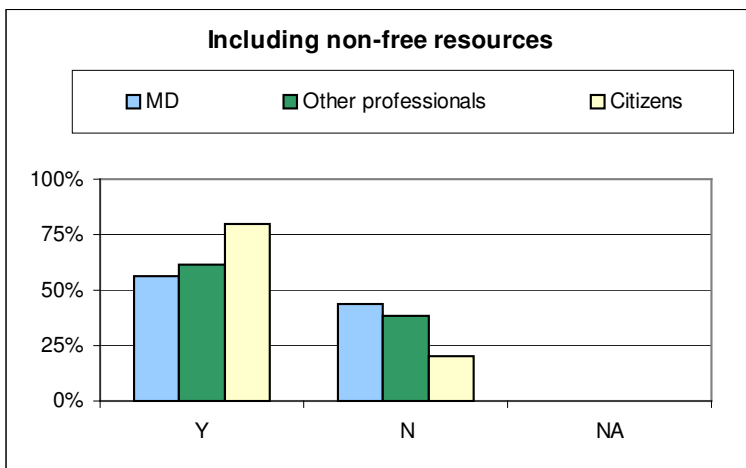


Figure 5, instead suggest that we radically change our selection policy about non-free resources.

In the beginning, indeed, we had decided to exclude all resources requiring an access fee not paid by anyone of the 3 institutions involved in the project.

About 60% of users, on the contrary, require to have in the catalogue also non-free resources.

To include them, however, we decided to follow these criteria:

- the resource should allow a free trial period, so that we can describe and evaluate it
- the cost should be acceptable for an individual user

3.1.2.3 Internet resources and search engines recommended by users

Table 4

Name of the resource	Address	Number of users who recommended it	Points *
Pubmed - Medline	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed	9	76
Cochrane Database of Systematic Reviews	http://www.cochrane.org e http://www.areas.it	9	71
IPASVI	http://www.ipasvi.it	5	38
OMIM	http://www.ncbi.nlm.nih.gov/Omim/;	3	30

	http://www.ncbi.nlm.nih.gov/		
Ministero della salute	http://www.ministerosalute.it	5	28
ISTAT	http://www.istat.it	3	19
OVID	http://medsifm001.medicina.unito.it/ovidweb.4.1.1/inizio2.cgi (local installation URL)	2	19
Istituto Superiore di Sanità (ISS)	http://www.iss.it/	2	17
GIMBE	http://www.gimbe.org/home.htm ; http://www.gimbe.it	2	16
Joanna Briggs	http://www.joannabriggs.edu.au	2	14
MURST	http://www.murst.it/	2	14
NORD	http://www.rarediseases.org	2	11
Associazione italiana Terapisti della riabilitazione/fisioterapisti	http://www.aitr.it	1	10
Virgilio	http://www.virgilio.it/home/index.html	1	10
BSPED	http://www.bsped.org.uk/	1	10
Medscape	http://www.medscape.com	1	10
Telethon Institute of Genetics and Medicine	http://www.tigem.it/	1	10
Medscape 's Diabetes and Endocrinology MedPulse(R)	http://www.medscape.com/diabetes-endocrinologyhome	1	10
Karolinska Institute	http://www.ki.se/	1	10
CDC PREVENTION GUIDELINE (US)	http://www.phppo.cdc.gov/cdcRecommends/AdvSearchV.asp	1	10
Amedeo.com	http://www.amedeo.com	1	10
cancernet	http://www.cancernetwork.com	1	10
Obgyn.net	http://www.obgyn.net/	1	10
American Academy of Physical Medicine and Rehabilitation	http://www.aapmr.org	1	9
SINPIA	http://www.sameint.it/sinpia/indice.htm	1	9
Endocrine society	http://www.endo-society.org/	1	9
Prenatal Risk Assesment	http://www.w-cpc.org/pregnancy/teratogen.html	1	9
Ensembl	http://www.ensembl.org/Homo-sapiens/	1	9
CDG	http://www.cdc.gov/health/	1	9
Wemove	http://www.wemove.org	1	9
WHO publications Maternal and child health	http://www.who.int/dsa/cat98/mat8.htm	1	9
cancer.gov	http://www.nci.nih.gov/	1	9
CINAHL	http://gateway1.ovid.com/ovidweb.cgi?T=JS&MODE=ovid&PAGE=main&D=nursing (local installation URL)	1	9
National guidelines clearinghouse	http://guidline.gov/index.asp	1	9
Nursing standard	http://www.nursin.nursing-standard.co.uk	2	8
Associazione italiana ortottisti e assistenti oftalmologia	http://www.aiorao.it	1	8
Medscape pediatrics	http://mp.medscape.com/pediatrics/home/	1	8
International Society of Psychosomatic Obstetrics and Gynaecology	http://www.ispog.org/?404	1	8

The Genome Database	http://www.gdb.org/	1	8
Micromedex	http://www.micromedex.com/	1	8
NLM	http://www.nlm.nih.gov/nlmhome.html	1	8
Sites and databases related to genetics, cytogenetics and oncology	http://www.infobiogen.fr/services/chromcancer/dblinks.html#GL	1	8
NICE (istituto di eccellenza clinica UK)	http://www.nice.org.uk/Cat.asp?pn=professional&cn=toplevel&ln=en	1	8
Neonatologia	http://www.neonatologia.it	1	8
AIMAC	http://www.aimac.it/	1	8
Evidence-Based Nursing	evidencebasednursing.com/cgi	1	8
Servizio Informazioni e Valutazione Ausili - Fondazione Don Gnocchi	http://www.siva.it	1	7
Pneumonet (sito italiano di pneumologia)	http://www.pneumonet.it/	1	7
Elsevier Science, Health Sciences Division	http://www.us.elsevierhealth.com/	1	7
FMFPC	http://utenti.lycos.it/fmfpc/unklab.html	1	7
Nature genome gateway	http://www.nature.com/genomics/	1	7
Uhrad.com	http://www.uhrad.com	1	7
SaPeRiDoc (regione Emilia-Romagna)	http://www.saperidoc.it/	1	7
areamedlab.it	http://www.areamedlab.it	1	7
Società Italiana Medicina Fisica e Riabilitativa	http://www.simfer.it	1	6
Facoltà di medicina e chirurgia	http://www.medicina.unito.it	1	6
The Lancet	http://www.thelancet.com/	1	6
CDER	http://fda.gov/cder/index.html	1	6
United Leukodystrophy Foundation (ULF)	http://www.ulf.org	1	6
ANDRIA (società italiana)	http://space.tin.it/salute/pipuzzi/andria.html	1	6
BMJ	http://www.bmj.com	1	6
Nursesarea	http://www.nursesarea.it	1	6
Office of Rare Diseases	http://rarediseases.info.nih.gov/ord/	1	5
E-neuro	http://www.e-neuro.it	1	5
MEDICAL DICTIONARY	http://cancerweb.ncl.ac.uk/cgi-bin/omd?action=Home&query=	1	5
Dica33	http://www.dica33.net/	1	5
Società italiana di ginecologia e ostetricia	http://www.sigo.it/	1	4
Orphanet	http://www.orphanet.infobiogen.fr	1	4
Birth choice (UK)	http://www.birthchoiceuk.com/	1	4
RCOG (UK)	http://www.rcog.org.uk/home.asp?PageID=3	1	3
Cardiologia	http://www.gitic.it	1	3
QUBIsoft	http://www.cardionet.it; http://www.gastronet.it; e altri (dedicati alla neurologia, urologia ecc....)	1	3
The Department of Neurology	http://www.neuro.wustl.edu/	1	2
Farmacovigilanza.net	http://www.farmacovigilanza.com	1	2
Parlamento	http://www.parlamento.it	1	2
Comune di Torino	http://www.comune.torino.it	1	1

Sito medici di medicina generale	http://www.simg.it	1	1
----------------------------------	---	---	---

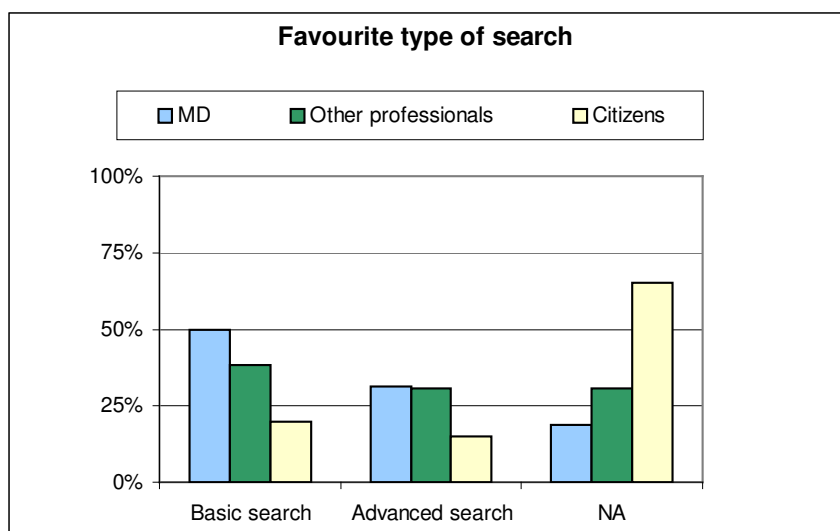
* Each user could recommend a maximum of 10 resources ranked by their importance. The first resource recommended was assigned 10 points, the second 9, and so on.

Quite obviously, web sites most used by our users are bibliographic databases and institutional sites. However, many resources of different kinds were recommended (sites about drugs, legislation, professional associations, EBM, etc.): cataloguers will analyze these resources in order to include them in the catalogue.

Answers about generic or specific search engines preferred by users, instead, are less interesting. Google and Virgilio (<<http://www.virgilio.it>>) are the most used general search engines, while specific biomedical engines are known only by very few users.

Figure 6 unveils an interesting aspect about search habits. Though many users (especially citizens) didn't answer, more beta-testers than we expected use advanced search in search engines (24%, while 35% is more comfortable with basic search and 41 doesn't answer). These data suggest that we grant special consideration to developing and improving this functionality in our project.

Figure 6

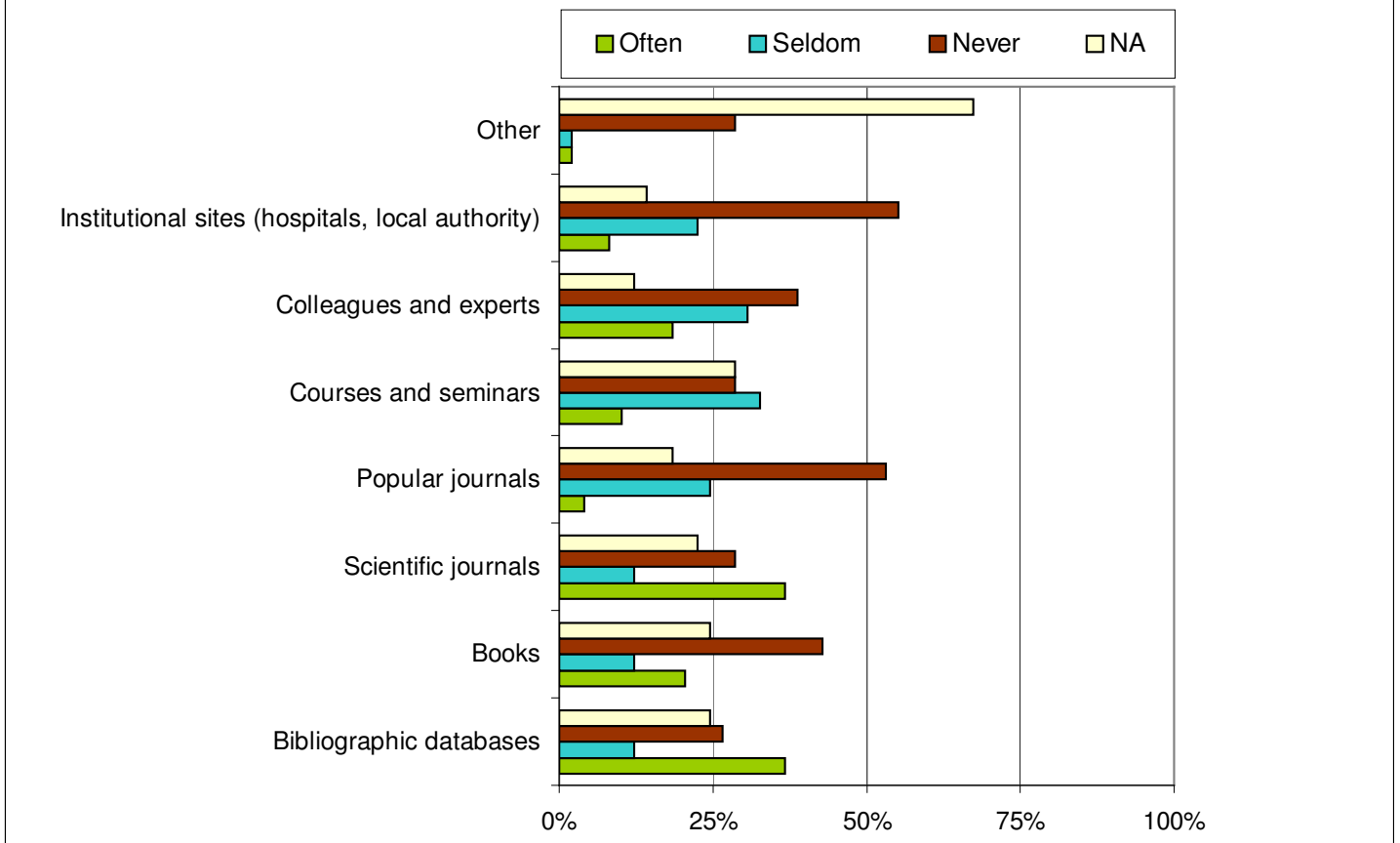


The last graph (Figure 7) shows that – a part from search engines and bibliographic databases – when users want to find useful Internet resources they very often ask to colleagues or friends, follow specific courses, or use more traditional sources such as books and journals.

It is highly positive that such a plurality of methods to gather information about the Internet exists, but at the same time data seem to suggest that building a central point to search for biomedical information could be very useful.

Figure 7

Users' favourite sources to find Internet biomedical resources



3.1.3 Users' evaluation of the beta version of ELISIR

The second questionnaire aimed to give beta-testers the possibility of trying ELISIR's search engine and give us suggestions and advice about how to improve it.

The test was performed from July 21 to August 7, 2002. Through ELISIR were available 191 records of catalogued resources. They were indexed using 84 MeSH terms in total. 134 of these resources were written in English, 63 in Italian, 9 in Spanish, 7 in French, 5 in German and 2 in Portuguese.

The following figures show the different search possibilities of the ELISIR search engine and an example of the help file.

Figure 8: basic search

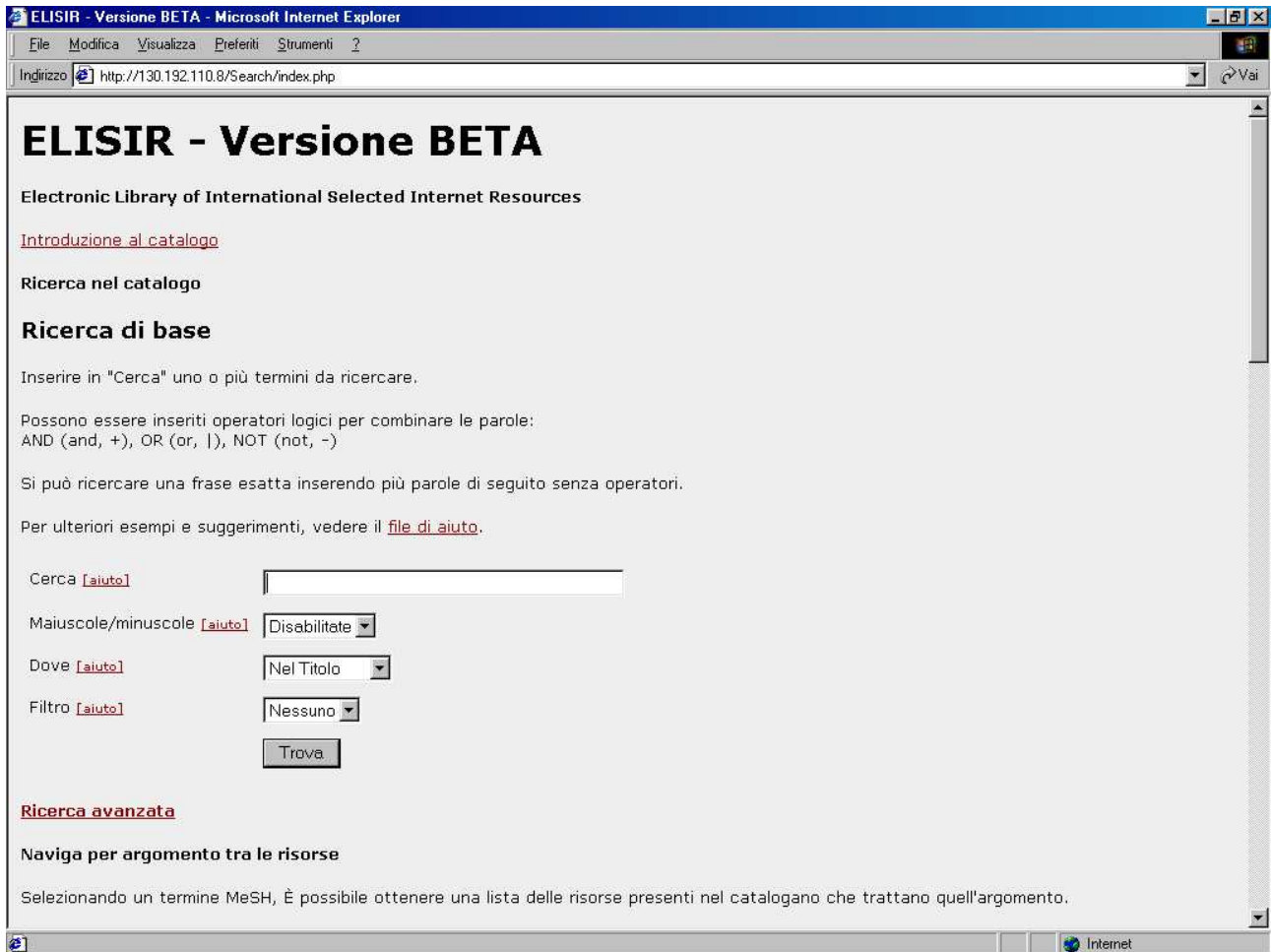


Figure 9: Advanced search

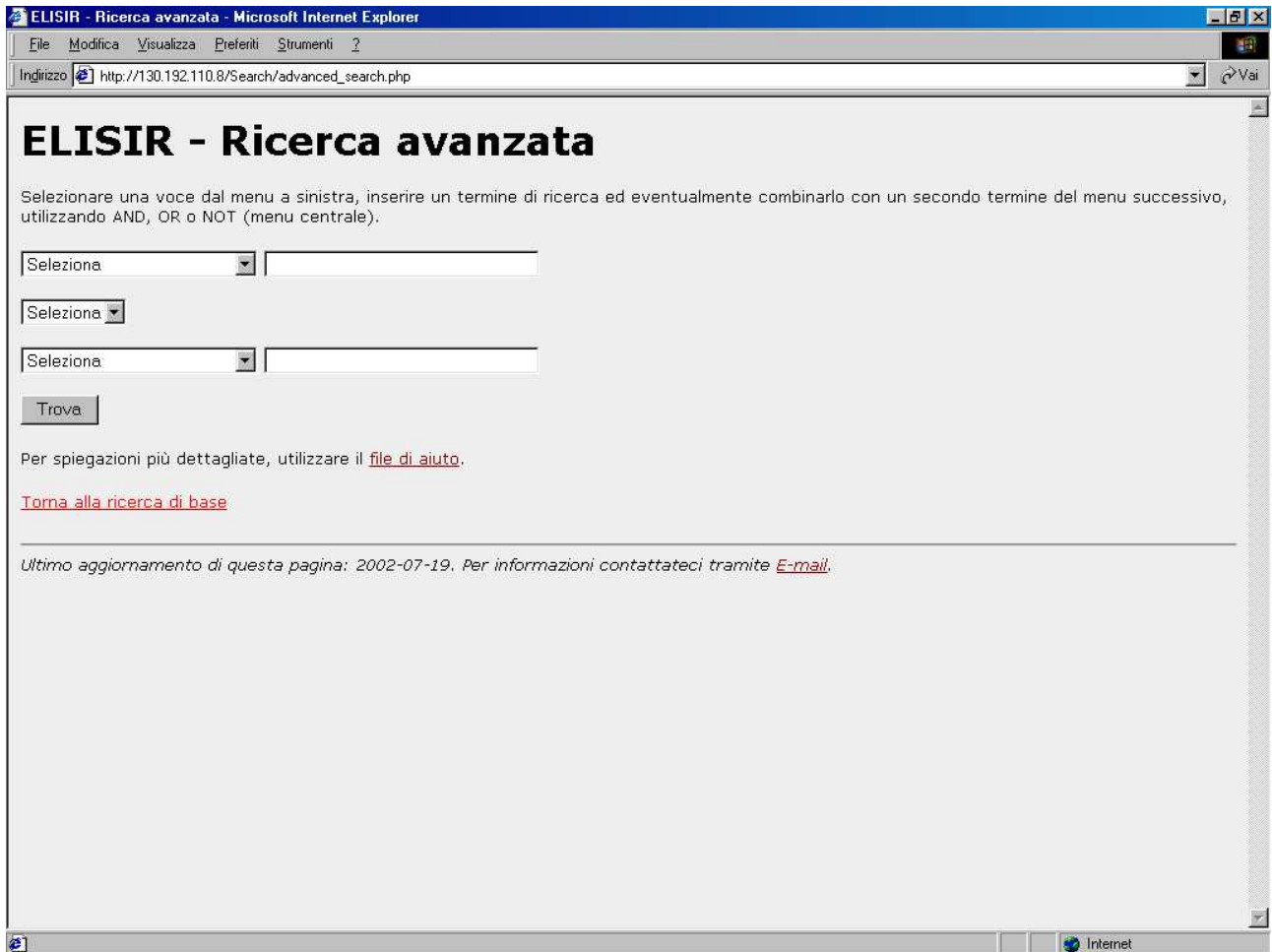


Figure 10: MeSH browsing

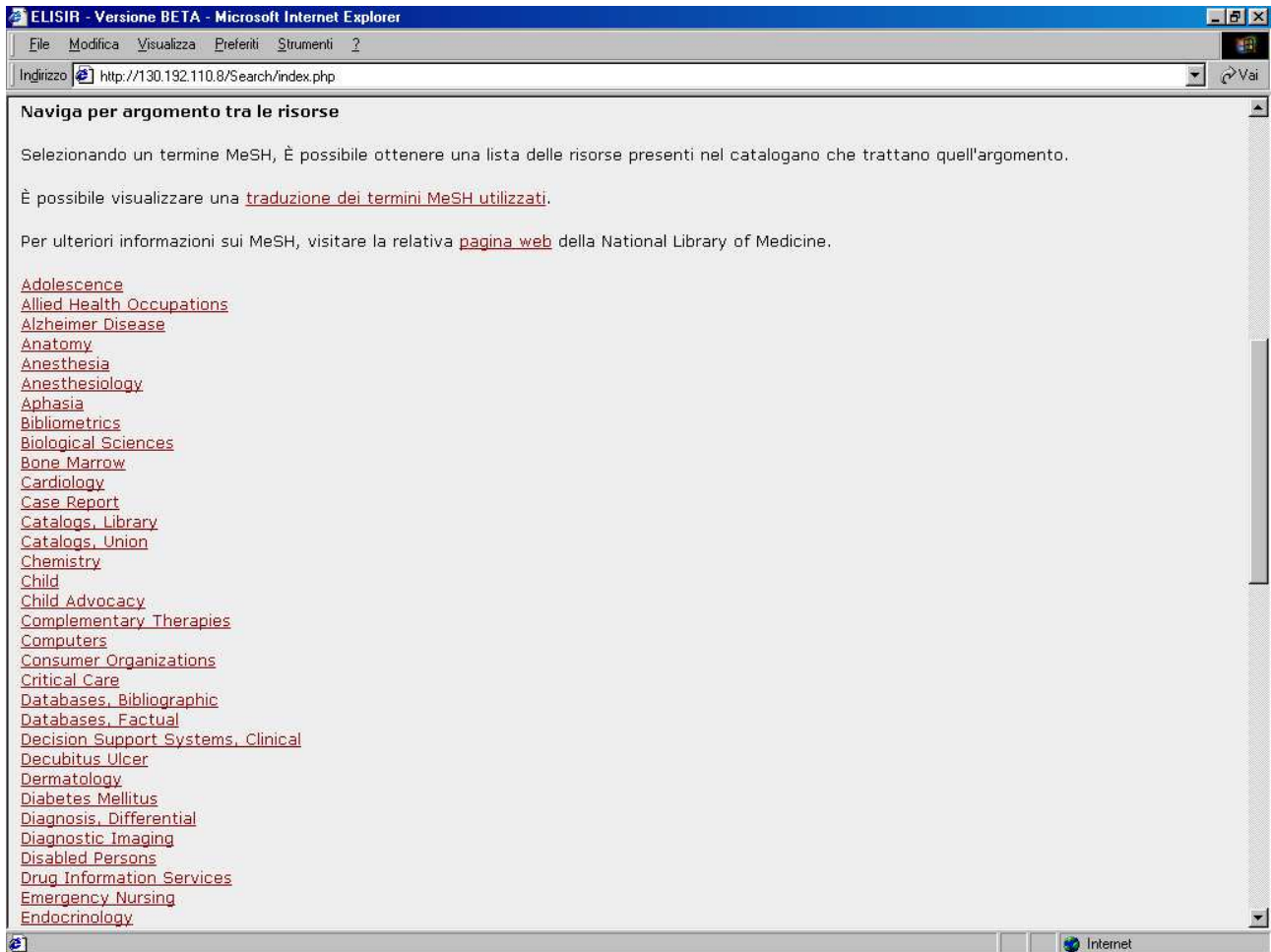
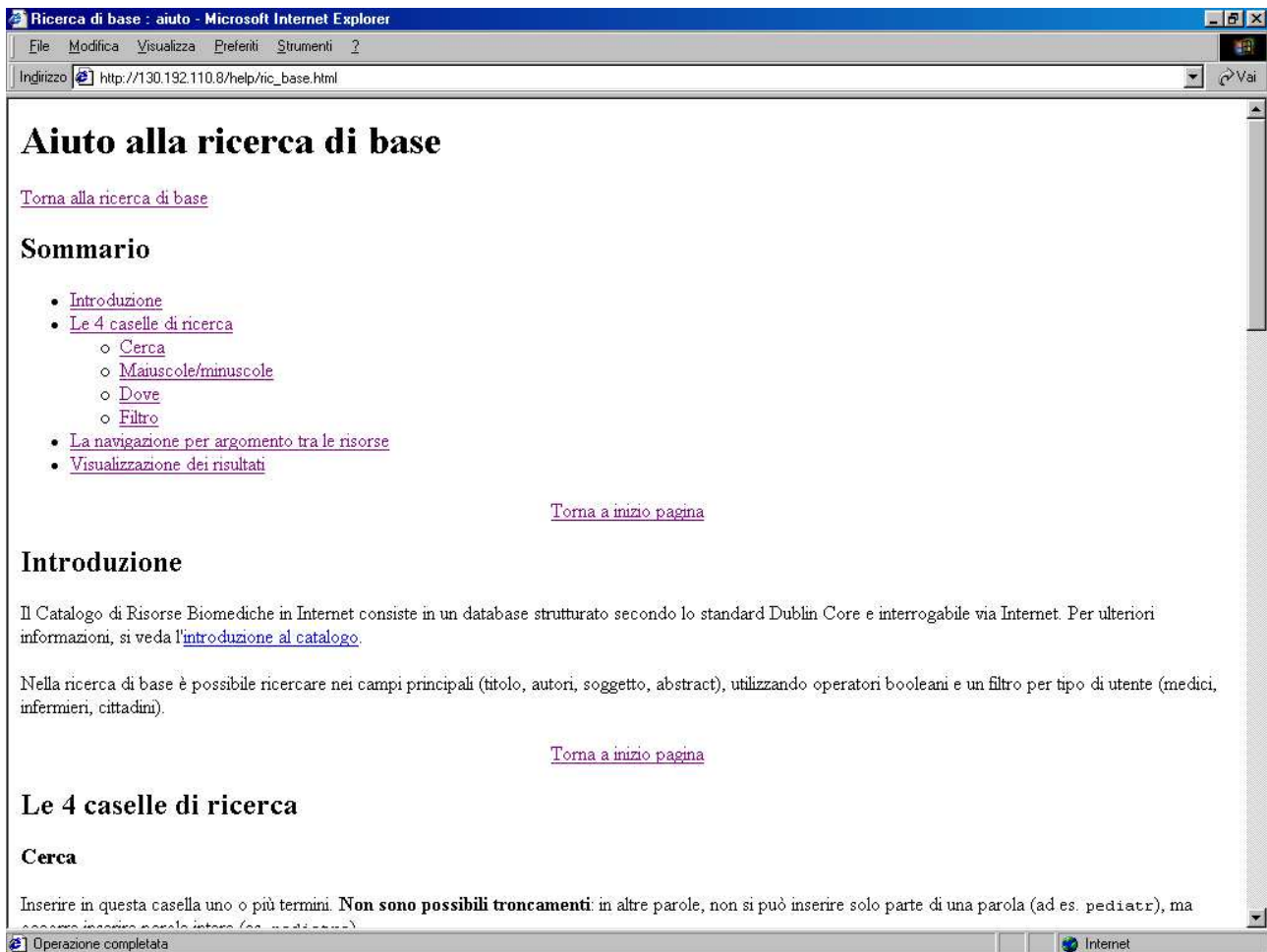


Figure 11: help file for basic search



The following graphs show the same data as for the first questionnaire: rates of answers and weight of the single categories on the total of answers.

Figure 12

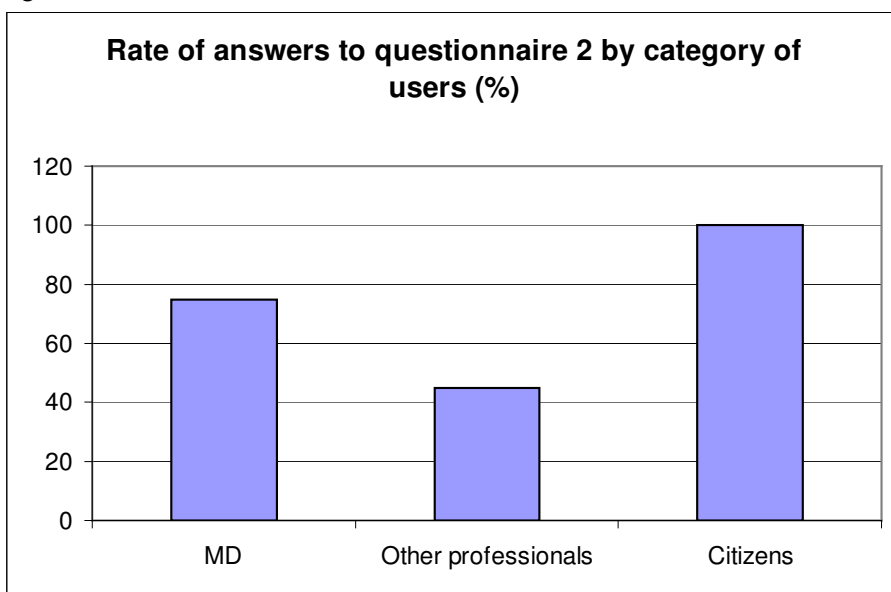
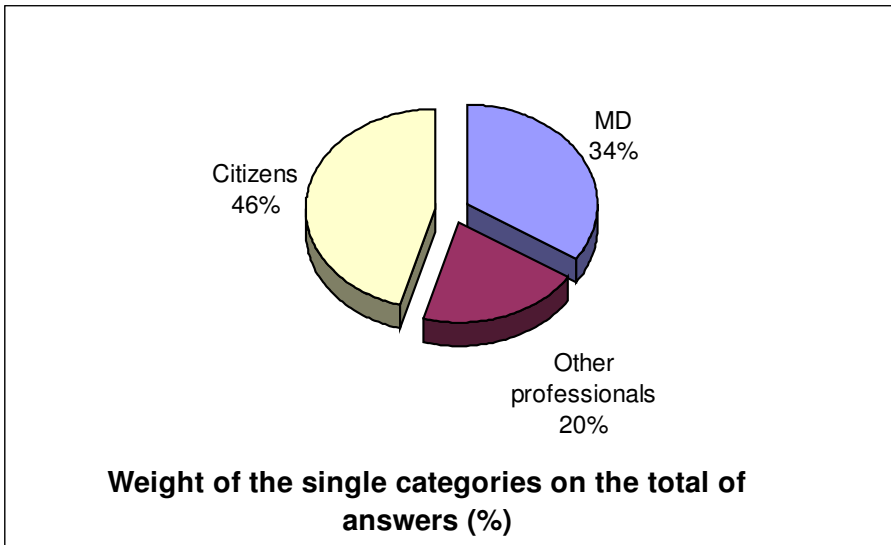


Figure 13



3.1.3.1 Technical issues

According to the data provided by Beta testers, they have used the catalogue for an average of 27' each.

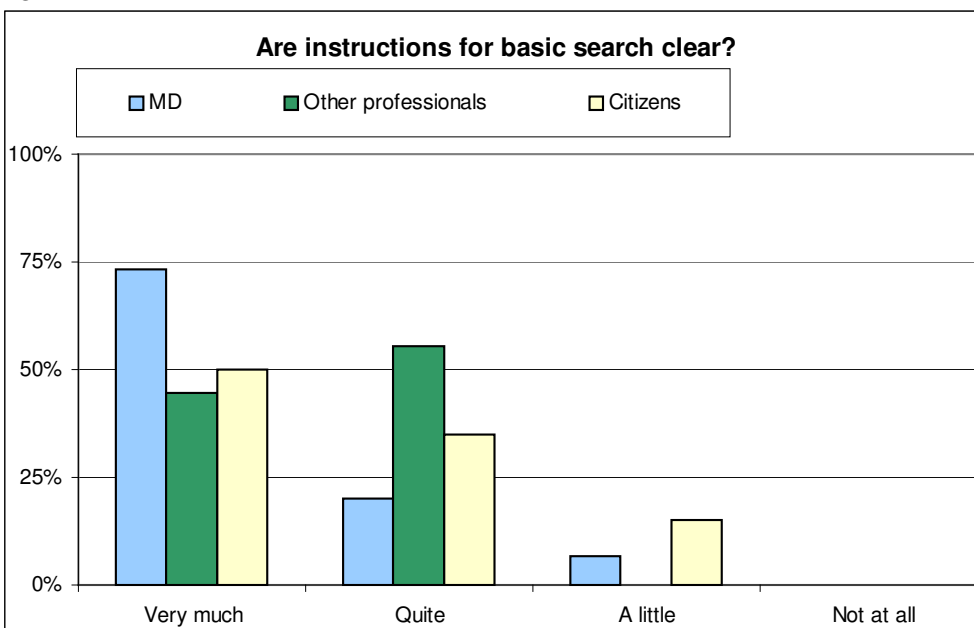
82% used Microsoft Internet Explorer (version 5, 5,5 o 6 or not specified), 14% Netscape Navigator (version 4.5 or not specified), 4% didn't answer.

The section of the questionnaire which aimed to point out possible technical problems showed a good performance of ELISIR.

The only serious problem was the slow answer to the queries (more than 15''), pointed out by 18% of users.

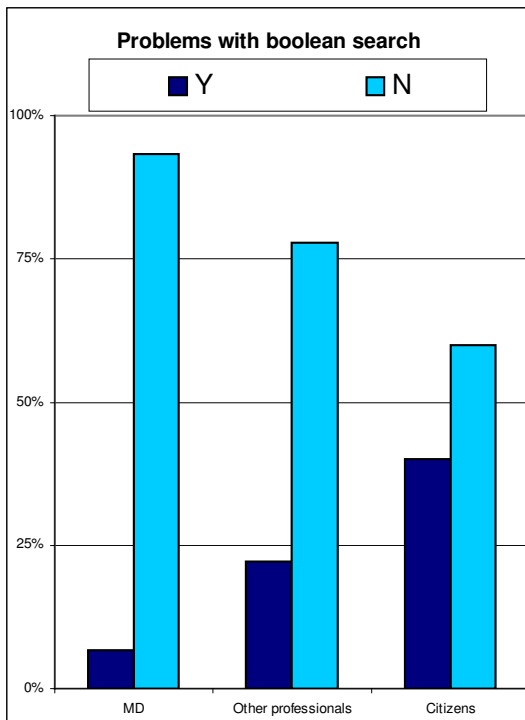
3.1.3.2 Usability of the user interface

Figure 14



Most users stated that search instructions on the basic search form were very or quite clear.

Figure 15

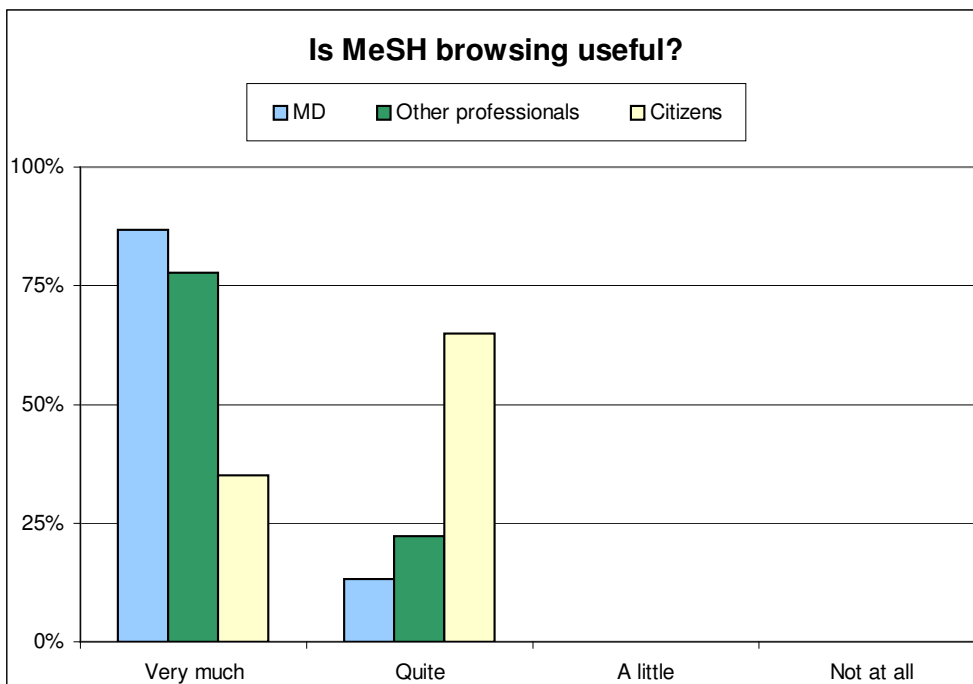


MDs and other professionals encountered in general only few problems with phrase and boolean searches. Citizens are, instead, less used to searching online: thus, they are less familiar with boolean operators.

Approximately 70% of users think that **advanced search** is “very useful”, while it is “quite useful” for 30%.

23% of beta-testers encountered problems with advanced search: they weren’t actually technical problems, however, but in most cases only an insufficient explanation of the search possibilities.

Figure 16



MDs and other professionals seem to appreciate particularly MeSH browsing: 47% and 56% of them, respectively, would prefer a more detailed subject indexing, though most of the total answers (77%) considers adequate the present subject indexing, which adopts the policy not to use more than 3 MeSH for each resource. For the majority of citizens MeSH browsing is only “quite useful”. This depends perhaps also on lower linguistic skills of citizens, as compared with the other two categories: many citizens, indeed, suggest that we translate the MeSH terms in Italian – what we have already decided to do (see paragraph 2.4).

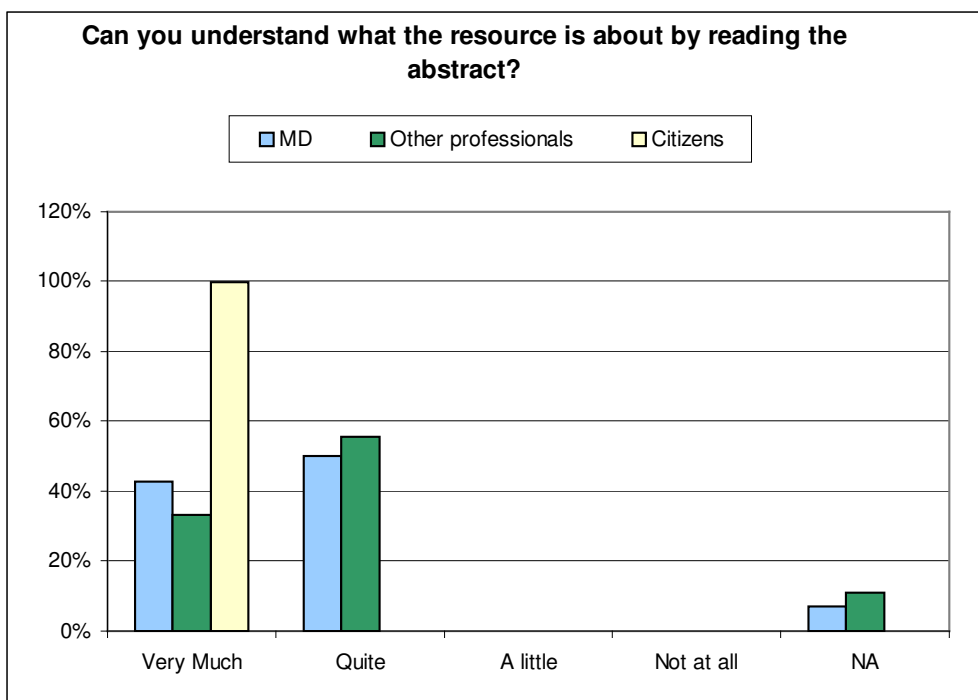
The present subject indexing was considered “very adequate” by 45% of users and “quite adequate” by 30%, while 25% didn’t answer.

We have taken into account the first issue by deciding to test a more detailed subject indexing for all resources: we decided to assign a maximum of 5 MeSH to a single resource, with the possibility to assign 10 MeSH terms to resources which were particularly rich of different section of content.

Finally, since less than 50% of beta-testers think that the present subject indexing is “very adequate” we should work on our indexing skills in order to improve them.

If tables of contents of the catalogued resources (where present) were perceived as “very important” by 59% of users and as “quite important” by 30%, the abstracts were very successful. 67% of users, indeed, manages to understand quite very what the resource is about by reading the abstract, (quite well for the 28%); MDs and other professionals, however, seem to appreciate abstract less than citizens (see figure 17).

Figure 17



20% of MDs and 44% of other professionals would even appreciate a more detailed abstract. However, since writing abstracts is one of the most time consuming activities, we will try in first instance only to guarantee a uniform minimal length in abstracts (about 6 lines).

Afterwards, we will be able to write more detailed abstracts if we can increase the number of human resources involved in the project.

Data types chosen to qualify DC element DATE are “very adequate” for 66% of our users and “quite adequate” for 20%.

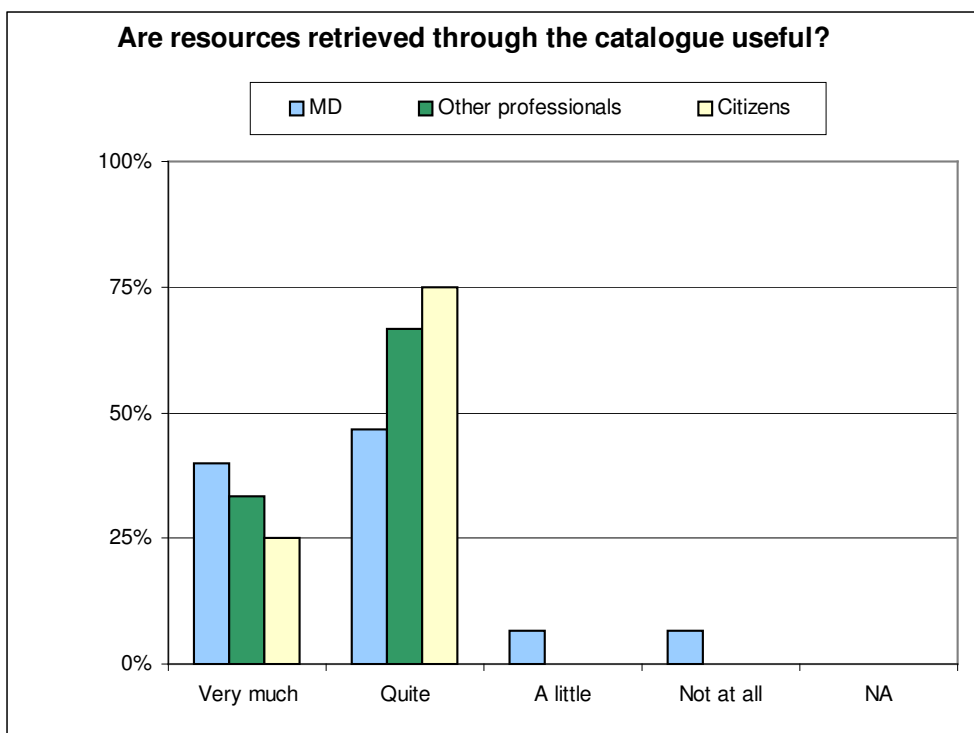
At present, it is possible to filter search results in ELISIR using 3 different user categories: MDs, nurses and citizens. 93% of beta-testers agree with this typologies, but 16% suggest that we include new categories, summarized below:

Table 5

Suggestion	Occurrences
Physical therapists	3
Dietists	2
Gynecologists	1
Ob-Gyn nurses	1
Pediatric nurses	1
Midwives	1
Pediatricians	1
Psychologists	1
Students	1

Since the filter can't be expanded to all these categories, we will only modify the resource by changing "nurses" in "other professionals".

Figure 18



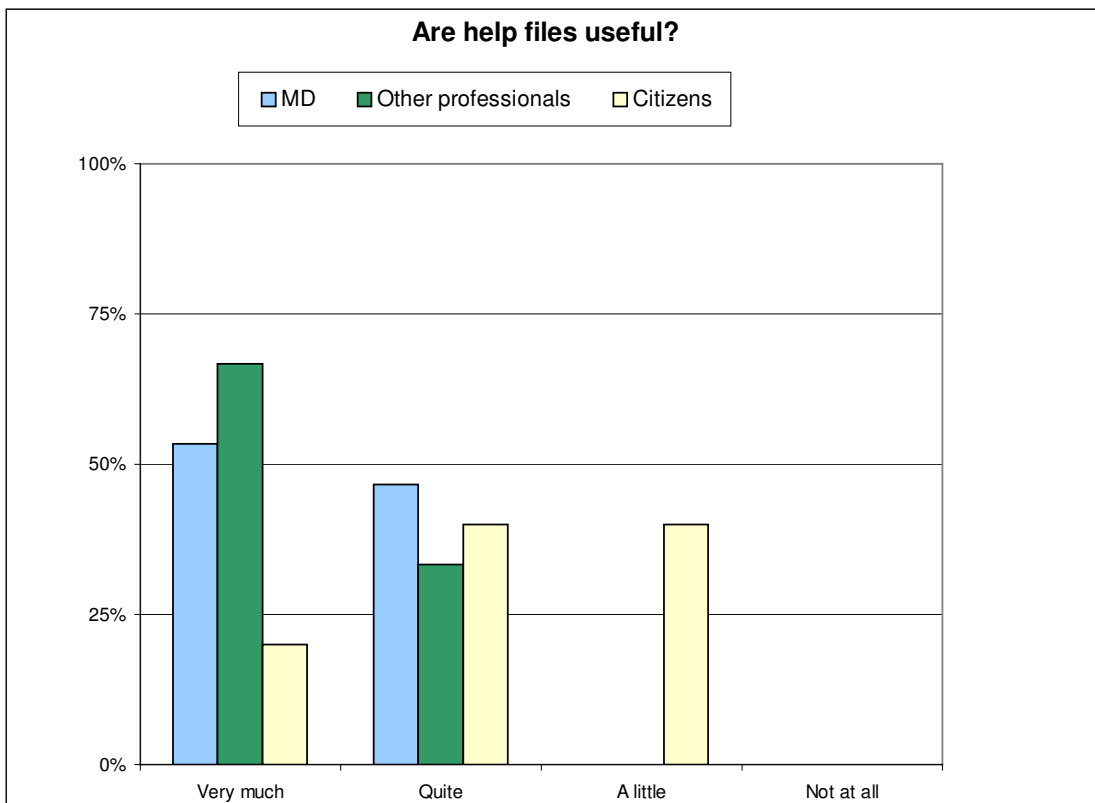
Even if the questionnaire was mainly about the usability of the search engine, we have also asked a very general question about the usefulness of the included resources.

Most of users (64%) considers them only "quite useful", while 32% "very useful" and 4% "a little useful" or "not useful at all".

We are confident that increasing the number of resources available, improving the skills of cataloguers and continuous users' feedback will help us improve the usefulness of ELISIR.

The present **number of displays** (brief: title, address and access restricted) is adequate for 98% of users, who wouldn't like to find different displays.

Figure 19



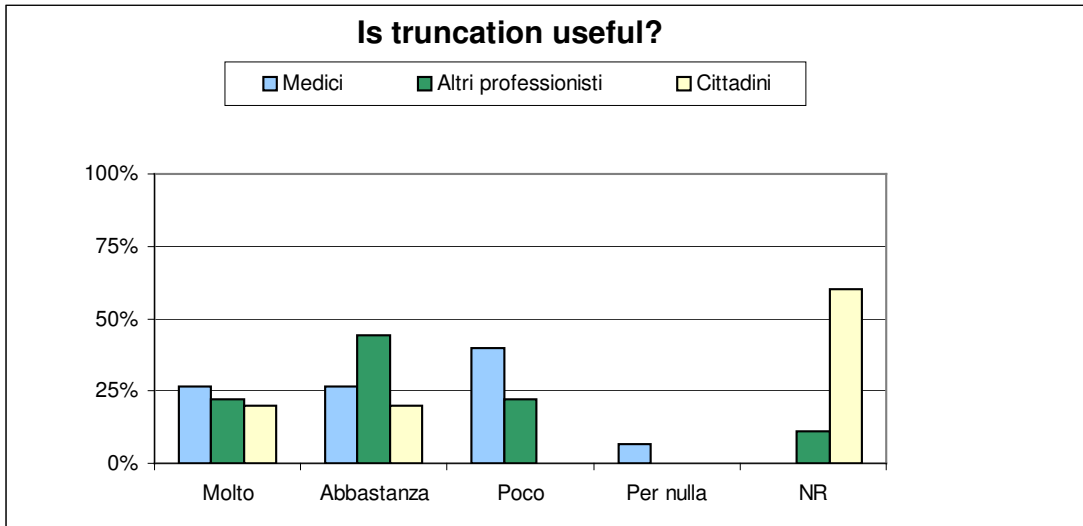
Most of users consider “very useful” (41%) or “quite useful” (41%) our help files. Many citizens, however, think that the language of the help file is not clear enough, and then consider the file only “a little” useful.

Most users considers our help files “very useful” (41%) or “quite useful” (41%). Many citizens, however (40%), consider the language of the file not so clear, and so the help file is considered only “a little useful”.

Thus, we decided to rewrite the more technical sections of the help files in such a way to make it more readable also for a non-specialized audience. Furthermore, we are planning to include in the files also graphics and links to web resource explaining search strategies in more details.

3.1.3.3 Users’ feedback about planned new functions of the search engine

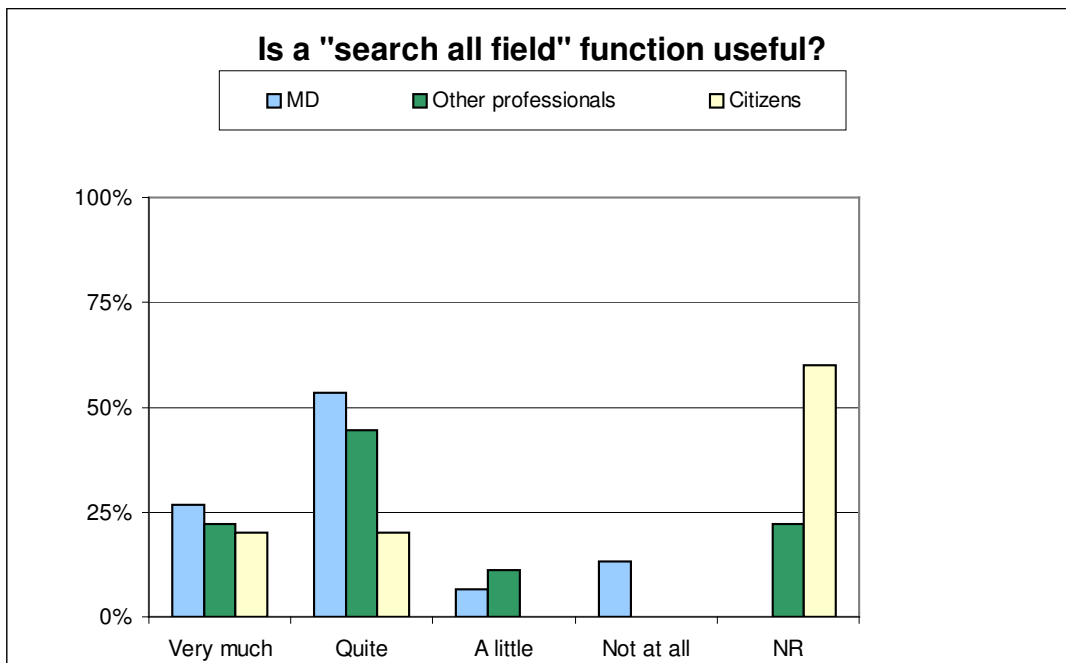
Figure 20



The current version of ELISIR's search engine doesn't allow truncation of search terms. User considered this functionality "very useful" (23%) or "quite useful" (27%), but 40% of MDs and 22% of citizens considers it only "a little useful". Many citizens (60%), perhaps not familiar with search strategies, didn't answer the question.

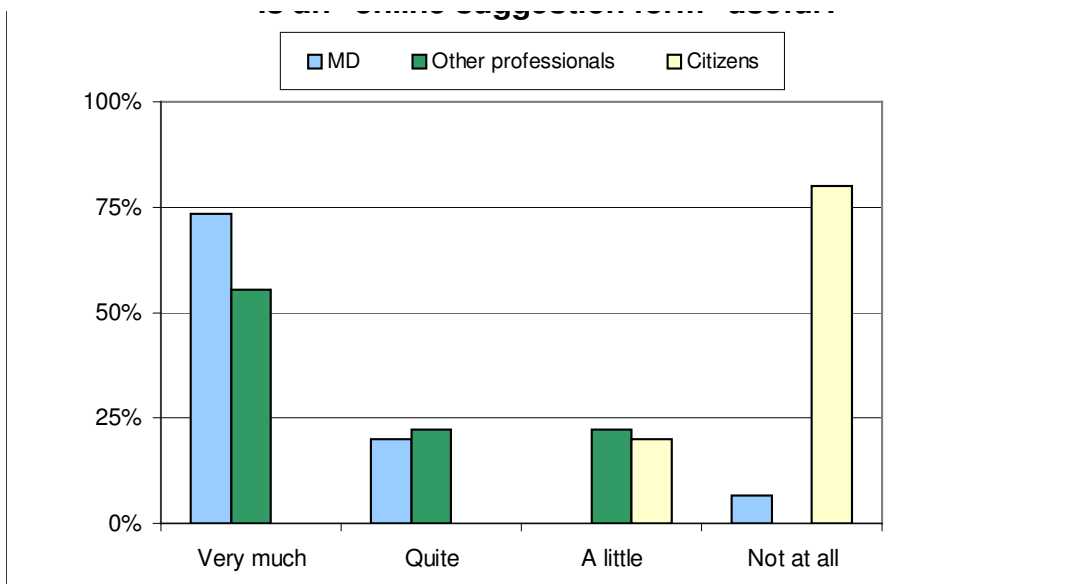
We will then implement this function in the new version, but we'll have to explain in detailed way how and when to use it and its advantages. We are also planning to monitor its use through the DBMS log files (see paragraph 4.2).

Figure 21



59% of users considers very much or quite useful this function, at present not supported.

Figure 22



In the final release of ELISIR's search engine we had planned to give users the possibility to recommend resources by an online form published on our web site.

MDs and, partially, other professionals strongly agree with this idea, while citizens doesn't seem interested.

Since implementing this form and the related database is not so expensive, we have decided to create it anyway. We'll keep in mind, though, that to collect citizens' feedback on this issue we'll have to use other methods (e.g. focus groups).

3.3 Continuous post-implementation feedback

We have seen that a big questionnaire can offer plenty of useful suggestions when planning and implementing a resource.

However, collecting and elaborating data is a very long and difficult task, while users' time is always a very scarce resource.

Thus, we decided to ask users to answer another similar questionnaire only in 12 months' time after ELISIR's final implementation. The goal will be to evaluate again the resource as a whole and to ask suggestions about improvements.

For continuous feedback, instead, we have chosen these two options:

- creating an **online form** by which users are encouraged to recommend resources to be included in the catalogue and to point out problems encountered. The form will be linked to a PostgreSQL database, which will allow cataloguers to view immediately user's suggestions and problems, to answer them rapidly and to adopt adequate solutions. Besides, data will be archived for future statistics.
- Creating 3 **focus groups**, one per category of users. Each will have 5 members e will meet twice a year. These groups will discuss efficiency of the resource, problems and amendments, both from an IT, and from a catalographic point of view. Reports from these working groups will be included in the technical reports of the project.

4. Measuring access and monitoring the system

Apart from direct feedback from users, discussed above, there other interesting possibilities of monitoring the usage of the resource which are offered by the web server and DBMS software.

Though far from being perfect⁷, these tools can help answering the following questions:

⁷ See Gardois 2001, especially the bibliography.

- Is the site used, and how much?
- Are there particular sections used less or more than others?
- Are there technical problems on the server and how frequently do they happen?
- Which are the most popular queries to the search engine?
- Consequently, how much and how are boolean search, truncation and phrase search used?
- From which external site do users come from when they visit our resource?

In the following paragraphs there are some notes about how we have used or we plan to use these tools for analyze indirect users feedback.

However, available data never allow to discover the identity of our users. Besides, complying with laws on privacy, we won't give any details about IP addresses of our users.

4.1 Usage statistics from Apache log file

A web server records every file request coming from the Internet.

Every request is recorded in a line of a text file called log file, saved in a directory of the web server itself.

Many kinds of log files are available, the most important being – in Apache - `error_log` (useful to track technical problems) and `access_log`.

We'll try to give an example of analysis of this second log file, which reports data similar to the ones contained in this line.

```
212.210.172.4 - - [22/Jul/2002:17:52:24 +0200] "GET
/Search/results.php?id_uri=124 HTTP/1.0" 200 2370
"http://130.192.110.8/Search/uri.php?id_mesh=49" "Mozilla/4.0
(compatible; MSIE 6.0; Windows 98)"
```

The line can be subdivided in the following sections, to understand it better:

```
212.210.172.4 - - : IP address of the user who has requested the page
[22/Jul/2002:17:52:24 +0200] : date and time of the request
"GET : HTTP method requiring the download of a page
/Search/results.php?id_uri=124 : URL of the requested page
HTTP/1.0" : protocol used
200 : code indicating that the download was successfully completed
2370 : number of bytes transferred
"http://130.192.110.8/Search/uri.php?id_mesh=49" : URL of the "referrer", or the
from which the user was directed to our page
"Mozilla/4.0 (compatible; MSIE 6.0; Windows 98)" : indicates user's browser (in
this case Microsoft Internet Explorer 6.0) and operating system (in this case Windows 98).
```

A log file is made of thousands or more lines like this.

To analyze it and obtain useful data it is necessary a kind of software called log analyzer. We have used, for this study, Mach 5 Enterprise Fast Stats 2.81 (<http://www.mach5.com/products/analyzer/index.php>).

The following data could be a useful complement to integrate data from questionnaires and are referred to the test phase of the ELISIR's search engine (July 21 to August 7, 2002).

4.2.1 General statistics

Firstly, the log file related to this period was analyzed, excluding all the request of files to directories other than the one of the search engine, and all the requests coming from IP addresses of cataloguers and robots⁸.

The following data are only a small part of the ones which can be obtained by an accurate analysis of the log file.

We intentionally restricted ourselves to data which can give complementary indications and can confirm or contradict the results of the analysis of the questionnaires.

A first set of data concerns general statistics about system usage (Table 6). "Visiting users" (probably in some cases, it is the same user connecting in different moments) were 114 and downloaded pages 804, with an average of about 9 page per user (but, if we consider that the users were indeed 44, the average increases to 18 pages per user). These data confirm that beta-testers used the resource intensively, since available pages were about 10 in total.

The average time users spent on our web site was about 16': very different from the 27' drawn from the questionnaire. The reasons are probably that users tend to overestimate time spent using a resource and that some users included in the total time also the time used in completing the questionnaire.

The good performance of the system was confirmed by the fact that there were no incomplete downloads of files.

Table 6

Item	Value
Average Hits per Day	56,8
Average Hits per User	8,9
Average time spent by a single user of the web site	15'45"
Hits on Pages	804
Incomplete downloads/file requests	0 (0%)
Total Visiting Users	114
Unique IP Addresses	59

4.2.2 Usage of single pages

In table 7 we show the data related to the usage of single parts of the search engine. Not surprisingly, the most downloaded page was the one containing search results (from basic or advanced search)

Since it was downloaded 222 times and there were no incomplete downloads, it reasonable to state that users have searched the database 222 (5 searches per user): another indirect indication that the system was tested accurately.

However, the detailed display page was downloaded only 112 times: users, the, often only viewed the brief display given in answer to a query, without going further to read the abstract or to visit the resource

Advanced search was used 69 times over the 222 times in which users have viewed results from queries, which means 31%. Thus, advanced search – though appreciated – was used noticeably less than the basic one.

The result related to help files, instead, is far more difficult to be interpreted: only 32 downloads for the help file for basic search, 18 for the MeSH help, 17 the intro to ELISIR and 12 the advanced

⁸ Robots are software agents which automatically index enormous quantities of web pages and save references in the search engines' databases.

search help file. In general, help files were viewed 79 times: 1,8 times per user in total, but single files were viewed only 0,3 to 0,7 times per user.

This could mean that help files were not so useful because it was already quite clear how to use the engine. Another more probable hypothesis, though, confirmed by data from questionnaires, is that help files weren't used so much because when a user opened the first couldn't clearly understand the content...

Obviously, these data could be further analyzed, and probably in the future new parameters will be included.

Two interesting examples could be the following:

- **frequency of usage of single MeSH terms** in MeSH browsing. Data can be obtained analyzing the complete URL of the request of resources associated to a MeSH term, which is given in the form: `http://130.192.110.8/Search/uri.php?id_mesh=73` . As we can see, to each MeSH term is associated an ID: thus, it is only necessary to calculate the frequency of an ID to calculate how many times the corresponding MeSH was used.
- the **referrer function**. As already mentioned, the log file allows to view the page from which a user is directed to another page, also if both pages are on the same server. This allows to study users' paths inside a given resource, especially when pages are interconnected in such a complex way to allow access to a page from many different other pages. Careful analysis of these data can help establish special "advised" paths through a resource, by restructuring the whole web site architecture.

Obviously, data from log files are only and always partial. All the cache memories located between the server and the user hide a relevant rate of contacts. A typical example? Every time you hit the back or forward button of your browser.

Therefore, log file analysis is useful mainly to give complementary indications on the usage of a resource. It should never be used alone: when an administrator needs to make important decisions, direct users feedback is always necessary.

Table 7

Page Name	Description	Hits
/Search/search.php	Brief display of basic search or advanced search results	222
/Search/uri.php	Brief display of results of MeSH browsing	142
/Search/index.php	Home page of the ELISIR engine	129
/Search/results.php	Detailed results of basic or advanced search or MeSH browsing	112
/Search/advanced_search.php	Advanced search page	69
/help/ric_base.html	Help file on basic search	32
/help/mesh_tr.html	Help file on MeSH browsing	18
/help/intro.html	Intro to ELISIR	17
/help/ric_av.html	Help file on advanced search	12

4.2.3 Technical aspects: web browsers and operating systems

Knowing what OS and browser our users prefer is useful to try to create compatible and easily viewable web pages. Anyway, the best practice is always writing simple and clear HTML conforming to the W3C's WAI (<<http://www.w3.org/WAI/>>).

Data about our users show that an overwhelming majority (more than 97%) use Microsoft as an operating system. Linux users are 2.6%.

Table 8

Operating System	Hits	Percentage
Windows 98	648	65.26%
Windows 2000	103	10.37%
Windows 95	100	10.07%
Windows ME	55	5.54%
Windows NT	35	3.52%
Linux	26	2.62%

As for browsers, the different versions of Microsoft Internet Explorer represent about 80% of the total. The rest is divided between 18% of Netscape Navigator and 2% of not specified browsers (probably running on Linux).

Table 9

Browser	Hits	Percentage
IE 5.x	602	60.62%
IE 6.x	167	16.82%
Navigator 4.x	155	15.61%
Other	21	2.11%
IE 4.x	20	2.01%
Navigator 3.x	16	1.61%
Navigator 5.x	10	1.01%
Navigator 2.x	2	0.20%

4.2 PostgreSQL log file

Another precious data source about the usage of the resource could be the log file of the DBMS used to archive the catalogue data.

From an analysis of an Apache log file, indeed, it is not possible to retrieve the content of query executed through the POST method of HTTP: we can see how many times a page has been loaded, but not what the user was searching for.

To solve this problem, PostgreSQL will be configured in order to produce a log file readable by a special log analyzer.

At present, we have planned to PostgreSQL Log Analyzer 1.1 (<http://www.samse.fr/GPL/pg_analyzer/README.html>). The test phase will begin the next fall, with the following objectives:

- discovering the most used keywords
- discovering the most used combinations of single terms
- monitoring truncation and phrase search
- discovering how many searches gave zero results and why

Analysis of these data should provide indications about what resources users would expect to find in ELISIR and about how efficiently users can search. Besides, these data could be useful to plan user education.

5. Conclusions

Planning and implementing a catalogue of Internet resources wouldn't be possible without a direct involvement of users in every step of the process.

Though precious, librarians' experience and skills are not enough to determine user needs and especially the right way to encounter them.

It is necessary to complement the more traditional tools used to assess users' feedback with new tools allowing a constant monitoring of the usage of the resource.

The carefully planned automation of some phases of the selection process, together with a clear idea of users' needs, can prove very useful in sparing cataloguers' time to improve the quality of a catalogue.

The adoption of standards widely adopted both about cataloguing rules and about software solutions could contribute to widen the community built upon this project, in the perspective to create a national community. To pursue of this last objective, however, it will be necessary a careful consideration of organizational models and of the financial and institutional aspects. Finally, adopting the open source solution for implementing and distributing specific software, could help to create a high quality system shared and used by the scientific community.

6. Bibliography

1. Aeree S, Mee-Kyung S. Evaluating health information sites on the Internet in Korea: a cross-sectional survey. *Asia Pac.J.Public Health* 2001 ;13 Suppl:S19-22.
2. Aguillo I. A new generation of tools for search, recovery and quality evaluation of World Wide Web medical resources. *J.Manag.Med.*2000;14(3-4):240-8.
3. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. *JAMA* 2001 May. 23-30;285(20):2612-21.
4. Bin L, Lun KC. The retrieval effectiveness of medical information on the web. *Int.J.Med.Inf.* 2001 Jul;62(2-3):155-63.
5. Boulos M, Roudsari A, Carson E. Towards a semantic medical Web: HealthCyberMap's tool for building an RDF metadata base of health information resources based on the Qualified Dublin Core Metadata Set. *Med.Sci.Monit.* 2002 Jul;8(7):MT124-36.
6. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. *Health Educ.Res.* 2001 Dec;16(6):671-92.
7. Cucina RJ, Shah MK, Berrios DC, Fagan LM. Empirical formulation of a generic query set for clinical information retrieval systems. *Medinfo.* 2001;10(Pt 1):181-5.
8. D'Alessandro D, Kingsley P. Creating a pediatric digital library for pediatric health care providers and families: using literature and data to define common pediatric problems. *J.Am.Med.Inform.Assoc.* 2002 Mar-Apr;9(2):161-70.
9. Darmoni SJ, Thirion B, Leroyt JP, Douyere M, Lacoste B, Godard C et al. A search tool based on 'encapsulated' MeSH thesaurus to retrieve quality health resources on the internet. *Med.Inform.Internet.Med.* 2001 Jul-Sep;26(3):165-78.
10. Darmoni SJ, Thirion B, Leroy JP, Douyere M. The use of Dublin Core metadata in a structured health resource guide on the internet. *Bull.Med.Libr.Assoc.* 2001 Jul;89(3):297-301.
11. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the Internet for medical information. *J.Gen.Intern.Med.* 2002 Mar;17(3):180-5.
12. Eysenbach G, Kohler C, Yihune G, Lampe K, Cross P, Brickley D. A metadata vocabulary for self- and third-party labeling of health web-sites: Health Information Disclosure, Description and Evaluation Language (HIDDEL). *Proc.AMIA.Symp.* 2001:169-73.
13. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA* 2002 May 22-29;287(20):2691-700.
14. Fasella, C. Il modello bibliografico FRBR e il Dublin Core. Seminario nazionale sui metadati. ICCU - Roma, 3 aprile 2001 / ICCU Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche. < <http://www.iccu.sbn.it/Fasella.doc> >
15. Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *BMJ* 2002 Mar 9;324(7337):569-73.
16. Gardois, P. Measuring the performance of a biomedical digital library: web site, e-journals and databases. 2000. <<http://pacs.unica.it/alghero2001/proceedings/015.htm>>
17. Gardois, P. Evaluating the quality of medical information on the Internet: a brief bibliography. 2001. < <http://www.icml.org/posters/post28/poster28.htm>>
18. Gillespie J. Web Page Design for Designers Web site. 2002. <<http://www.wpdfd.com/>>

19. Joubert M, Aymard S, Fieschi D, Fieschi M. Quality criteria and access characteristics of Web sites: proposal for the design of a health Internet directory. Proc.AMIA.Symp. 1999:824-8.
20. Joubert M, Aymard S, Fieschi D, Fieschi M. ARIANE: a mediation framework with health information sources. Medinfo. 2001;10(Pt 1):343-7.
21. Kinzie MB, Cohn WF, Julian MF, Knaus WA. A user-centered model for web site design: needs assessment, user interface design, and rapid prototyping. J.Am.Med.Inform.Assoc. 2002 Jul-Aug;9(4): 320-30.
22. Kunst H, Groot D, Latthe PM, Latthe M, Khan KS. Accuracy of information on apparently credible websites: survey of five common health topics. BMJ 2002 Mar 9;324(7337):581-2.
23. Kushniruk AW, Patel C, Patel VL, Cimino JJ. 'Teleevaluation' of clinical information systems: an integrative approach to assessing Web-based systems. Int.J.Med.Inf. 2001 Apr;61(1):45-70.
24. Malet G, Munoz F, Appleyard R, Hersh W. A model for enhancing Internet medical document retrieval with "medical core metadata". J Am Med Inform Assoc. 1999 Mar-Apr;6(2):163-72.
25. Monopoli, Maria and Nicholas, David. A user evaluation of Subject Based Information Gateways: case study ADAM. Aslib proceedings 2001;53(1):39-52.
26. Nielsen, J. useit.com : Jakob Nielsen's Website. 2002. <<http://www.useit.com/>>
27. Norman F. Organizing medical networked information (OMNI).Med Inform (Lond). 1998 Jan-Mar;23(1):43-51.
28. Ozok AA, Salvendy G. Measuring consistency of web page design and its effects on performance and satisfaction. Ergonomics. 2000 Apr;43(4):443-60.
29. Pandolfini C, Bonati M. Follow up of quality of public oriented health information on the world wide web: systematic re-evaluation. BMJ. 2002 Mar 9;324(7337):582-3.
30. Price SL, Hersh WR. Filtering Web pages for quality indicators: an empirical approach to finding high quality consumer health information on the World Wide Web. Proc AMIA Symp. 1999:911-5.
31. Purcell GP, Wilson P, Delamothe T. The quality of health information on the internet. BMJ. 2002 Mar 9;324(7337):557-8.
32. Randolph AG, Kane J. Potential for meeting clinician information requirements in an intensive care unit (ICU) via the web. Proc AMIA Symp. 1998:96-100.
33. Risk A, Petersen C. Health information on the internet: quality issues and international initiatives. JAMA. 2002 May 22-29;287(20):2713-5.
34. Risk A, Dzenowagis J. Review of internet health information quality initiatives. J Med Internet Res. 2001 Oct-Dec;3(4):E28.
35. Rodgers RP. Searching for biomedical information on the World Wide Web.J Med Pract Manage. 2000 May-Jun;15(6):306-13.
36. Rosenfeld L. and Morville, P. Information Architecture for the World Wide Web : Designing Large-scale Web Sites. Cambridge, O'reilly, 1998. 224 p.
37. Sekikawa A, Aaron DJ, Acosta B, Sa E, LaPorte RE. Does the perception of downloading speed influence the evaluation of web-based lectures? Public Health. 2001 Mar;115(2):152-6.
38. Shon J, Musen MA. The low availability of metadata elements for evaluating the quality of medical information on the World Wide Web. Proc AMIA Symp. 1999:945-9.
39. Wilson P. How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the internet. BMJ. 2002 Mar 9;324(7337):598-602.

Credits

The authors wish to thank C. Tortorelli and T. Di Giacomantonio for their valuable help.