**S. J. Darmoni[1], J. P. Leroy[1],
F. Baudic[2], M. Douyère[1],
J. Piot[2], B. Thirion[2]**

[1]Computer and Networks Department
and [2]Medical Library, Rouen
University Hospital, Rouen, France

CISMeF: A Structured Health Resource Guide

Abstract: In 1999, the Internet has become a major source of health information. The objective of CISMeF is to catalogue and index the main French-speaking health resources. In September 1999, the number of indexed resources totaled over 7,100 with a mean of 75 new sites per week. CISMeF uses two standard tools for organizing information: the Medline bibliographic database MeSH thesaurus and the Dublin Core metadata format. Resources included in CISMeF are described by the following: title, author or creator, subject and keywords, description, publisher, date, resource type, format, identifier, and language. To index resources, CISMeF uses five levels of hierarchy: "meta-term", category, keyword, subheading, and resource type. CISMeF contains a thematic index, including medical specialities and an alphabetic index. CISMeF respects the Net Scoring, criteria to assess the quality of health information on the Internet. The CISMeF project offers a valuable tool for the French-speaking health community: 2,500 computer users visit the Web site each working day.

Keywords: Abstracting, Indexing, Catalogue, Internet, Vocabulary

1. Introduction

In 1999, the Internet can be considered as a major source of scientific and health information [1]. For the healthcare professional and the health consumer, the access of accurate information on the Internet is not easy; therefore, there is a profusion of directories and search engines available in this new media [2]. However, directories, such as Nomade [http://www.nomade.fr/], Carrefour [http://www.carrefour.net/] or Yahoo [http://www.yahoo.com], or search engines, such as Altavista [http://www.altavista.com] do not permit the end-user to obtain a clear and organized presentation of the available health information. This limits the use of these tools. There is no catalogue available of resources in a specific medical speciality, e.g., in neurology. These directories or search engines contain a large number of sites but the organization and the hierarchy of the data is not adapted to health and medicine. In this field, there is a specific need for a specialized classification. The up and down consultation of the tree in this classification permits the acquisition of greater information.

Furthermore, the need of a rigorous human classification is mandatory to coherently organize heterogeneous resources, such as patient associations, electronic journals, mailing lists, and clinical guidelines. In a previous study, we demonstrated that manual catalogues vs. search engines were less sensitive but far more specific [3]. The main problem for the end-user is, therefore, to find useful information. Automated indexing has other drawbacks: difficulty in indexing non-text media, problems with word indexing (contact, syntax, morphology, content, polysemy, synonyms and granularity), and lack of structured information (keywords, controlled thesaurus, metadata e.g., resource types).

The objective of CISMeF is to assist health professionals during their search of electronic information and knowledge available on the Internet. CISMeF is a project that was initiated by the Rouen University Hospital (RUH). Its Universal Resource Locator (URL) is http://www.chu-rouen.fr/cismef (Fig. 1).

CISMeF began in February 1995 with the creation of the RUH's Web site. Its URL address is http://www.chu-rouen. fr (S.J.D. and B.T are the two co-webmasters).

The scope of CISMeF covers healthcare disciplines and medical sciences. This means that besides physicians who were historically the first target, other health professionals, such as nurses, midwives, veterinarians, physiotherapists, nutritionists, will find resources devoted to their professions. More recently, patients and the general public have found valuable and reliable health information specially written for them.

This resource guide is necessary because (a) there is an extensive amount of information potentially accessible for the health professional; (b) it is often difficult to easily separate the information for the health professional from patient information; and above all (c) the absolute requirement in medicine is to know the source and the quality of the information available on the Internet.

2. Material and Methods

2.1 Hardware and Software

CISMeF was implemented in February 1995 on a Sun running Sun Unix. We are planning to change the machine and the operating system in October 1999 from Unix to Lynux. CISMeF is a Web site currently using the NCSA freeware http server. The new server will be an Apache http. The Getstats, version 1.2 and Webtrends Log Analyzer, version 4.5.2. programs evaluate the use of the Web page after excluding requests by our Hospital Information System (1,800 computers).

2.2 Standards

CISMeF is entirely based on static HTML. We had originally used the HTML 2.0 standard, in order to be readable by the vast majority of Web browsers. We are now using the newer HTML 3.2 standard and planning to use XML when creating the CISMeF database. CISMeF uses two standard tools for organizing information: the MeSH (Medical Subject Heading) thesaurus from the Medline bibliographic database [4] (US National Library of Medicine) and the Dublin Core metadata format [5].

2.3 Methodology and Realization of the Catalogue

2.3.1 Why Indexing only French Health Resources

In November 1994, when the RUH was first connected to the Internet, we rapidly observed the absence of a specific catalogue of French-speaking health resources. In contrast, several very good catalogues for English-speaking resources such as DDRT (Diseases, Disorders and Related Topics [http://www.mic.ki.se/Diseases/index.html], Medical Library and Medical Information Center, Karolinska Institute, Stockholm, Sweden) or MedWeb, Health Science Center Library of Emory University-Us [http://www.medweb.emory.edu/MedWeb/] were already in operation (see also our selection of these catalogues in the following URL: http://www.chu-rouen.fr/ssm/listemed.html).
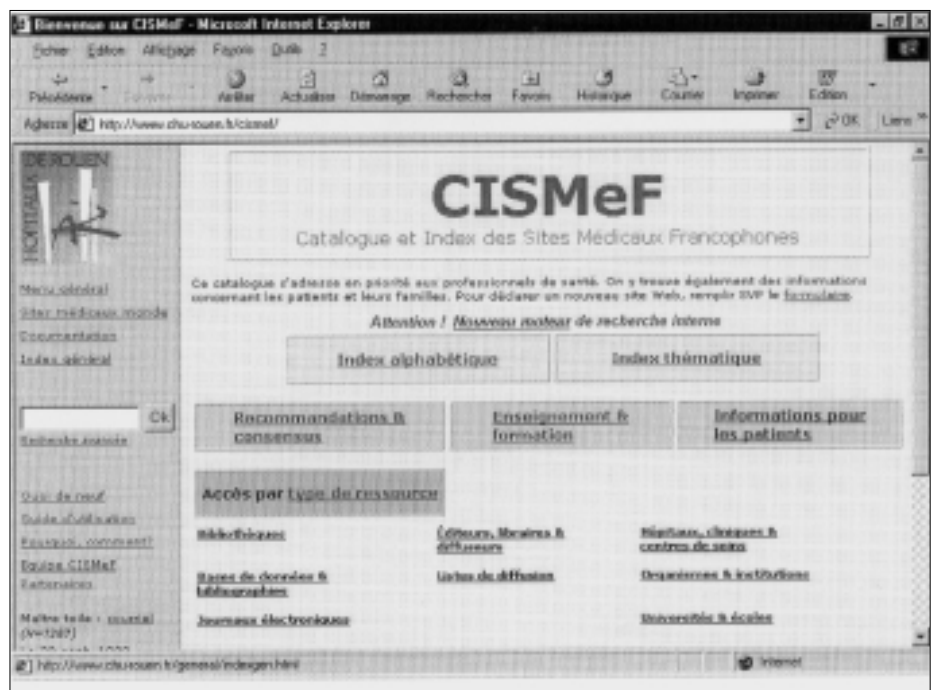

Fig. 1 CISMeF homepage screenshot.

Therefore, since its creation, CISMeF has only catalogued and indexed French-speaking health resources, independent of their origin.

The CISMeF method entails a four-fold process: resource collection, filtering, description and index. One deputy medical librarian (J.P.) performs the resource collection and the information watch. The editorial boards filters and selects the resources. Two deputy medical librarians (F.B. and M.D.) describe and index resources. The chief medical librarian (B.T.) is a 'super-indexer' in charge of checking the indexing. There is a 30-minute meeting daily with the medical informaticians (S.J.D. and/or J-P.L.) for double-checking.

2.3.2 Resource Collection

The resource collection is performed on a daily basis. French-speaking directories and search engines are checked, such as Carrefour, Ecila, Eureka, Francite, Nomade, Toile du Quebec, especially their "what's new" pages. The information watch and resource collection are also performed on a monthly visit to the French NIC (Network Information Center) to check new sites in the ".fr" domain (see http://www.chu-rouen.fr/documed/docum.html#VEILLE). A total of 576 health webmasters (8% of the sites included in CISMeF) have sent us an Email or a specific form to be indexed in CISMeF. Indexing priority is given to Internet sites of institutions and scientific societies. Resources include sites and high quality documents, issued especially from evidence-based medicine, practice guidelines, consensus development conferences, teaching and education resources.

2.3.3 Filtering and Selection

In order to include only reliable resources, CISMeF uses the Net Scoring [6], which consists of 48 criteria to assess the quality of health information on the Internet. These criteria are grouped in eight categories: credibility, content, links, design, interactivity, quantitative aspects, ethics, and accessibility. Some of the criteria have been inspired by a US white paper [7]. The description of a site should permit the evaluation of the quality of its content. Some resources are not introduced in CISMeF if they do not respect basic, particularly ethical, criteria.

Since December 1996, CISMeF has created an editorial board, based on the policy of the hospital managers to control its development, quality, validity,

reliability, consistency and coherence. This editorial board has developed a graphic chart (each HTML document must include the logo and the name of the hospital, the name and the E-mail of the author and the dating of the material). The editorial board peer reviews each resource included in CISMeF. If necessary, we use external peer reviewing. In 1999, we are planning to create a scientific board to optimize the peer review process.

2.3.4 Description and Indexing

Cataloguing a site is necessary because it helps the end-user to estimate, in advance, the type of information and to evaluate its content. This process also saves time for the end-user. Resources included in CISMeF are described by the following elements from the Dublin core project [http://purl.org/DC/about/element_set.htm]: title, author or creator, subject and keywords, description, publisher, date, resource type, format, identifier, and language. The following fields are specific to CISMeF: institution, city, province or state, country, target, free access, sponsorship.

We also include the following elements from the Dublin core metadata format in the metadata of CISMeF's HTML documents: language, keywords and their synonyms in English and in French, qualifier and resource types. For example: <meta name=«DC.language« content=«fre«> <meta name= «DC.subject.keywords« content= «(SCHEME=MeSH) euthanasie; euthanasia«> <meta name=«DC.type« CONTENT=«(SCHEME=CISMeF) texte.répertoire«>.

We have organized our catalogue with the MeSH thesaurus [4], which contains 19,232 terms in its 1999 version and nine levels of classification. This thesaurus is precise, rigorous and annually updated. We also use the French translation of this thesaurus [http://dicdoc.kb.inserm.fr:2010/basismesh99/mesh.html], performed by the French Medlars Center, the National Institute for Health and Medical Research (INSERM, and more specifically the DIC-DOC Network). In some rare cases, one resource cannot be perfectly indexed using the MeSH thesaurus: we are using "manual mapping" to search

the nearest MeSH term, e.g. a resource coping with dysmelia has been indexed with the MeSH term ectromelia.

In CISMeF, each keyword is "de facto" a MeSH Major Topic with a mean of 1.4 MeSH terms per resource. MeSH subheading permits a focus on a subfield of a MeSH term, e.g., chloride/toxicity. We also use a French translation of the MeSH subheadings [http://dicdoc.kb.inserm.fr:2010/basismesh99/ind_fr_eng.html], which is less systematically used in CISMeF compared to Medline. CISMeF resource type is a generalization of the publication type of Medline. We have added types which are specific to the resources available

Table 1 CISMeF resource types controlled list.

- advertisements [PT] *
- architectural drawings [PT]
- commercial company
- community networks
- database [PT]
  - database, bibliographic
- directory
  - annual directory
  - catalogs [PT]
  - registry
  - resource guides [PT]
- education
  - teaching material
    - educational courses
    - instruction [PT]
    - problems and exercises [PT]
    - tutorial
  - teaching structure
    - school
    - university
  - training
- establishment, institution, organization
- foundation
- hospital
  - hospital department
- image database
- library
- museum
- newsgroup and discussion list
- patient information
- periodicals [PT]
- publisher
- research structure
- scientific society
- search tools
- society
- software
- text
  - bibliography [PT]
  - congresses [PT]
  - consensus development conference [PT]
  - dictionary [PT]
  - dissertation, memoir
  - educational courses
  - encyclopedias [PT]
  - guide
  - guidelines
    - practice guidelines
  - journal article [PT]
  - legislation [PT]
  - medical thesis
  - monograph [PT]
  - problems and exercises [PT]
  - technical report [PT]
- trade association, trade society

* [PT] stands for Medline publication types

on the Internet, such as association, patient information, community networks. The controlled list of resource types is given in Table 1. Resource type describes the nature of the resource and MeSH describes the subject of the resource. For example, in case of a clinical guideline about carbon monoxide intoxication, 'carbon monoxide poisoning' is the MeSH keyword and 'clinical guidelines' is the resource type. In CISMeF, each description also contains the geographic localization, including the city, the province or state, and the country of origin.

Example of a description of a document indexed in CISMeF:

Adolescent pregnancy [http://www.cps.ca/english/statements/AM/am94-02.htm].

Adolescent Medicine Committee, Canadian Paediatric Society (CPS). In: Canadian Journal of Paediatrics 1994; 1 (2): 58-60 [prevention, diagnosis, management, counselling, bibliography]. -Ca

keywords: adolescence; pregnancy in adolescence resource type: clinical practice guideline

2.3.5 Structure of the CISMeF Catalogue

CISMeF contains a thematic index, including medical specialties [http://www.chu-rouen.fr/ssf/santspe.html] and an alphabetic index [http://www.chu-rouen.fr/ssf/santpath.html] (see Figure 1: CISMeF homepage screenshot). Both indexes use the Medline thesaurus. A brief description of each site indexed in CISMeF is systematically added.

To index the Health resources CISMeF uses five levels of hierarchy: (1) CISMeF "meta-term" (2) MeSH category (3) MeSH keyword (4) MeSH subheading (5) CISMeF resource type. The levels 1 and 5 are specific for CISMeF. The levels 2, 3 and 4 use the Medline database structure.

The CISMeF "meta-term" [8] is generally a medical speciality or a biological science, e.g., cardiology or bacteriology. For example, on the cardiology page, the sites of general interest about this speciality are indexed and described, using the MeSH keyword cardiology, followed by a list of starting

points of related categories and other associated MeSH terms. The categories are Cardiovascular agents, Cardiovascular diseases, Cardiovascular physiology, Cardiovascular system, Cardiovascular surgical procedures, and Diagnostic techniques, cardiovascular. The MeSH terms associated are Cardiac care, facilities, Cardiology service, hospital, and heart valve prosthesis. In CISMeF, each MeSH term (level 3 of the CISMeF structure) corresponds to a HTML document, which is organized first with MeSH subheading (level 4), then for each subheading, the resource types (level 5) (see Fig. 2: CISMeF mercury resources screenshot).

The alphabetic index uses the MeSH terms in English and their French translation, which permits bilingual search. The thematic index contains 96 terms, including 38 meta-terms, 25 MeSH categories and 33 MeSH keywords which in France are medical specialties, such as Aerospace Medicine. In addition, a general index is also available which contains MeSH synonyms in French and allows permuted utilization. Currently, we have indexed 7,112 resources with 1,819 MeSH terms. A mean of 75 sites and documents are indexed each week and approximately 30 new MeSH terms are added at the same time.

3. Results

CISMeF is efficient and an end-user-friendly solution to find French-speaking worldwide health resources on the Internet: 70% of these resources are located in France, 16% are from Canada, in particular the Quebec Province, 4% from Switzerland and Belgium, and 3% from Africa.

This Web site is principally and initially oriented for the health professional, although the general public may also have access to it. Many sites are devoted to both. There are no HTML documents with restricted access in the CISMeF Web site. Thus our traditional "end-users" are now not only healthcare practitioners but also patients, their families and anyone seeking health information [8]. Training should take into consideration the information needs of the lay person as well as those of the medical professional. Training
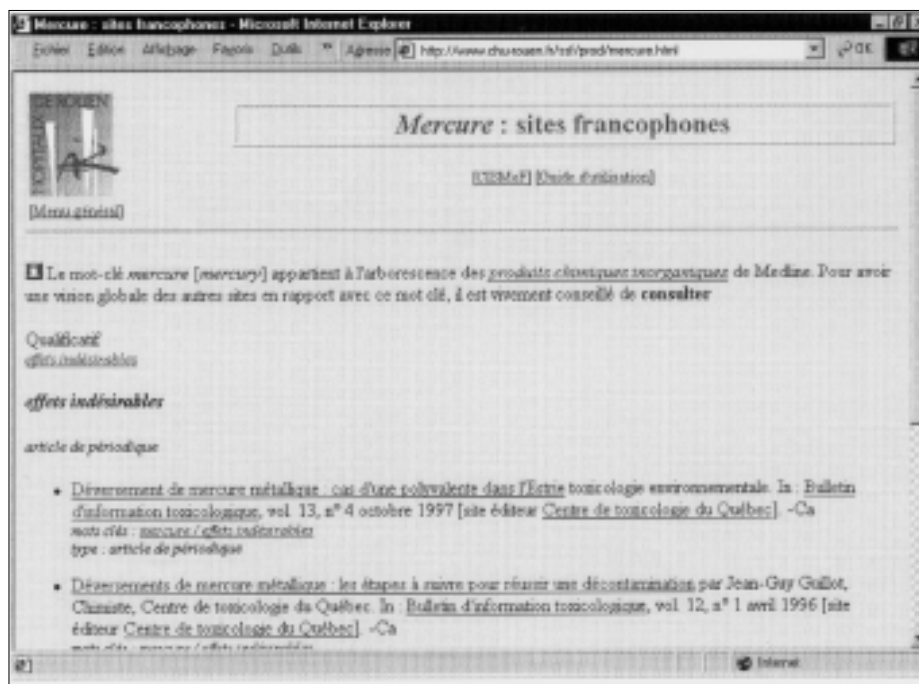

Fig. 2 CISMeF mercury page screenshot.

sessions on the use of our catalogue have been offered to Patients' Associations, especially to handicapped people at RUH since February 1999.

CISMeF has three priority axes: evidence-based medicine, teaching, and patient information. CISMeF also includes a list of clinical guidelines and consensus development conferences, hospitals, medical universities, health institutions, medical libraries, medical publishers, electronic journals, electronic textbooks, databases, teaching and CME, mailing lists, research laboratories and institutes, pharmaceutical firms, health and patient associations, and commercial companies in the health sector.

Since February 1995, some new features have been added to optimize the navigability and the access to the information for the end-user: (a) use of an internal search engine (full-text search), (b) a general index, and (c) a "what's new" page to easily display the newly indexed sites on a weekly basis. Since January 1997 it has also included an archive of the what's new pages. Two guides to use CISMeF are also on line, one for basic search and one for advanced search. CISMeF is accessible by the lowest common denominator of current browser technology.

3.1 Use Patterns of the Web Site

Use of the Web site increased in an approximate linear progression over time starting in February 1995. Our Web server software, which provides documents to users on request, does not know the identities of individual users, such as E-mail; the only identifying data available are the Internet IP addresses of the machines from which the users connect to the site.

Analysis of a representative period, the month of March 1999, showed that every working day approximately 2,500 machines visited our site (excluding ours). During the entire month, users from 68,601 computers made 280,195 requests for HTML documents from this Web site originating from 114 different countries (33.58% from France, 23.48% from USA, 8.03% from Canada, 3.33% from Belgium, 2.5% from Switzerland and only 0.19% from Africa). The geographical location of the connecting machine could not be determined in the remaining 23.89% of document retrievals (Internet IP addresses without domain name). These statistics underestimate the real figures due to the practice of file caching.

In March 1999, this Web site contained 13,605 hyper-links (including

3,687 internal hyper-links), 3,790 files (1965 HTML, 1112 GIF, and 250 JPG), and 218 megabytes of disk space. That is why it is of utmost importance to regularly check their validity. A software checks every hyper-link of the CISMeF site each month. Then the CISMeF team manually updates the broken links. The network traffic generated by the different Internet protocols (http, smtp, and nntp) represent between 0.3 and 1.5% (from 100 to 300 Mb/j) of the overall network traffic at RUH.

We also use the following indicator (Web impact factor [9]) to measure the current impact and potential ongoing future usage of CISMeF: number of sites, which have at least one hyperlink to our site [http://www.chu-rouen.fr/ dsii/html/pointeur.html]. Currently, our Web impact factor is over 800 (Altavista indicates more than 1,400 pages after exclusion of our internal links), including the most prestigious Health resource catalogues. Approximately 120 press articles released information about our Web site [http://www.chu-rouen.fr/dsii/html/presse.html]. In November 1997, in March 1998 and in November 1998, three surveys from the CESIM (Centre d'Etudes sur le Support de l'Information Medicale) showed that the RUH's Web site was the most frequently used by general practitioners (GP) with respectively 10%, 17.7% and 11.5% of their votes. In the same poll, only 11% of the GPs have an Internet connection, we can estimate that the RUH's web site is currently used by 1.2% of the French GPs. In March 1998, CISMeF has obtained the label "Experimentation of public interest" by an inter-ministry committee (procedure "Information Highways").

4. Discussion

The Internet facilitates the communication between health professionals and with the general public, and also improves the information access. However, most of medical resources available on the Internet only have a "marketing dimension" (description of the institution), and only a minority have a valid information content.

Several tools in the retrieval of health information on the Internet have been distinguished and structured:
– level 1: search engine, generalist or more specialized ones, such as Med-Hunt [http://www.hon.ch/];
– level 2: catalogue and index without thesaurus, such as Medical Matrix, US [http://www.medmatrix.org/];
– level 3: catalogue and index with thesaurus, such as the UMLS (Unified Medical Language System) metathesaurus [10] and MeSH thesaurus. The latter thesaurus is used in the following Health catalogues, Med-WebPlus-Us [http://www.medweb plus.com/], and DDRT;
– level 4: catalogue and index with thesaurus, metadata, and description of sites. To our knowledge, CISMeF and three other catalogues have now reached level 4: CliniWeb [11] [http://www.ohsu.edu/cliniweb/], Oregon Health Sciences University-Us, OMNI (Organizing Medical Networked Information-Uk) [http://omni.ac.uk/] [12] and HON (Health on the Net-Ch) [13].

OMNI indexes approximately 4,500 resources, mostly from UK, CISMeF about 7,000, mostly from France, Med-Hunt and HON around 40,000. OMNI and MedWebPlus are also using the UMLS metathesaurus to provide a conceptual network to the subject headings. OMNI, HON, and CliniWeb have also developed a structured database (dynamic HTML) which permits better searches. OMNI and CISMeF are using the Dublin Core metadata format, which is expected to become the dominant metadata format for Internet resource description [12].

It is difficult, especially for students, patients and the general public to evaluate the quality of the medical Web sites, which, in a majority of cases, are not peer-reviewed.

One main objective of CISMeF is to promote best medical practice and teaching. Therefore, we index high-quality documents available on the Internet on a priority basis. The organization of the CISMeF data model permits the discovery "by chance" of other "neighbor" sites and documents: e.g., a search about hemiplegia may permit the discovery of sites about paraplegia (relation of proximity) or more general-

ly going up in the tree about paralysis (relation of hierarchy). It also possible to find more information using the "see also" relation. It allows links between medical terms, which are not in the same tree, e.g., in the page about pain, see also terminal care. This "see also" relation seems difficult to automatically generate because it is based on human knowledge and not on a statistical model. This relation may be asymmetrical. If this relation is symmetrical between terminal care and pain, this is not the case for pain and bioethics. This link is only significant from pain to bioethics and not the opposite

Further challenges that CISMeF needs to address in the next months are to expand sites and high-quality documents, especially patient information, and to collaborate more closely with similar services, particularly in Europe (DDRT, HON and OMNI).

The major drawback of CISMeF is its weak technical level: CISMeF uses only static HTML and does not yet use a database which will allow more complex searches, e.g. guidelines in hepatitis, combining the explode command for keywords, qualifiers but also for metaterms and resource types. We plan to build this CISMeF database in the first semester 2000.

The major feature of CISMeF is its information structure model. CISMeF conceptually encapsulates the MeSH structure (category, keyword, qualifier) by adding two levels: one on top of it (metaterm) and one below it (resource type). This model is completely generic. It can help to design equivalent health catalogues in various languages where the MeSH thesaurus is already translated in these languages.

One key success of CISMeF is the typology of its webmasters: one medical librarian and one medical informatician. The CISMeF model was designed by these two individuals illustrating the synergy between these two professions. According to this experience, we suggest the use of this webmaster typology to design health web sites.

We did not formally and directly assess how end-users use CISMeF. During CISMeF training sessions, we observed that patients mostly use the internal search engine to access health information. Few end-users employ the

CISMeF model to perform better searches: in September 1999, 7% of the pages loaded are coping with MeSH categories or trees. In the near future, we will measure its real usefulness for different communities (physicians, nurses, and patients) in the different French-speaking countries by a questionnaire based on the Net Scoring. Some indirect elements indicate the CISMeF success: CESIM surveys, CISMeF use patterns and its Web impact factor. In order to enhance its quality, CISMeF respects the Net Scoring, e.g. its two webmasters personally answer each request.

CISMeF is a part of a wider project at RUH: digital library [1] and virtual university. We have already developed some parts of this digital library: access to Medline and 45 electronic full-text journals on Intranet (OVID provider) [13]. We plan to extend this library, giving access to electronic textbooks (e.g., the Harrison on the Internet). Our project of virtual university is to develop specific tools for students: a bank of multiple choice questions and a bank for standardized clinical examinations.

5. Conclusion

To help healthcare professionals and health consumers to more easily locate high-quality health information on the Internet, catalogues must use standard tools to describe and index resources.

REFERENCES

1. Schatz BR. Information Retrieval in Digital Libraries: Bringing Search to the Net. Science 1997; 275: 327-34.
2. Flannery MR. Cataloging Internet resources. Bull Med Libr Assoc 1995; 83: 211-5.
3. Darmoni SJ, Thirion B. Indexing the Web? A comparative study of three medical Web servers on the Internet: Cliniweb, "Diseases, Disorders and Related Topics". Omni. In: *Proceeding of the 1st European Congress of the Internet in Medicine.* 1996: 5-6. Unpublished. Available from Internet: <http://www.mednet.org.uk/mednet/dl4.htm>.
4. National Library of Medicine. Fact Sheet Medline. 6 July 1998 [Web document, accessed 11 Jan 1999] Available from Internet: <http://www.nlm.nih.gov/pubs/factsheets/medline.html>.
5. Weibel S, Juha H. DC-5: The Helsinki Metadata Workshop; A Report on the Workshop and Subsequent Developments. D-Lib Magazine 1998 February. Available from Internet: <http://www.dlib.org/dlib/february98/02weibel.html>.
6. Centrale Santé. Net Scoring: critères de qualité de l'information de santé sur l'Internet 20 Apr 1998 [Web document, accessed 27 Apr 1999]. Available from Internet: <http://www.chu-rouen.fr/dsii/publi/critqualv2.html>.
7. Ambre J, Guard R, Perveiler FM, Renner J, Rippen H. Health Information Technology Institute. Working Draft White Paper: Criteria for Assessing the Quality of Health Information 8 Apr 1999. [Web document, accessed 27 Apr 1999]. Available from Internet: <http://hitiweb.mitetretek.org/hswg/>.
8. Thirion B, Darmoni SJ. A simplified access to "MeSH Tree Structures" for Health professionals and patients, our new "end-users" on the Internet. Bull Med Libr Assoc 1999; 87: 480.
9. Ingwersen P. The calculation of WEB impact factor. J Documentation 1998; 54: 236-43.
10. National Library of Medicine. Fact Sheet UMLS Metathesaurus. 12 Aug 1998 [Web document, accessed 27 Apr 1999]. Available from Internet: <http://www.nlm.nih.gov/pubs/factsheets/online_indexing_system.html>
11. Hersh WR. Brown KE. Donohoe LC. Campbell EM. Horacek AE. CliniWeb: managing clinical information on the World Wide Web. J Am Med Inform Assoc 1996; 3: 273-80.
12. Norman F. Organising Medical Networks' information (OMNI). Med Inf 1998; 23: 43-51.
13. Boyer C, Baujard O, Baujard V, Aurel S, Selby M, Appel RD. Health On the Net: automated database of Health and medical information. Int J Med Inform 1997; 47: 27-9.

Address of the authors:
S. J. Darmoni,
Computer and Networks Department,
Rouen University Hospital,
1 Rue de Germont,
76031 Rouen Cédex,
France
E-mail: stefan.darmoni@chu-rouen.fr