MEDLINE*plus*: building and maintaining the National Library of Medicine's consumer health Web service

By Naomi Miller, M.L.S.
Systems Librarian

Eve-Marie Lacroix, M.S.
Chief, Public Services Division

Joyce E. B. Backus, M.S.L.S.
Systems Librarian

National Library of Medicine
8600 Rockville Pike
Bethesda, Maryland 20894

MEDLINE*plus* is a Web-based consumer health information resource, made available by the National Library of Medicine (NLM). MEDLINE*plus* has been designed to provide consumers with a well-organized, selective Web site facilitating access to reliable full-text health information. In addition to full-text resources, MEDLINE*plus* directs consumers to dictionaries, organizations, directories, libraries, and clearinghouses for answers to health questions. For each health topic, MEDLINE*plus* includes a preformulated MEDLINE search created by librarians. The site has been designed to match consumer language to medical terminology. NLM has used advances in database and Web technologies to build and maintain MEDLINE*plus*, allowing health sciences librarians to contribute remotely to the resource. This article describes the development and implementation of MEDLINE*plus*, its supporting technology, and plans for future development.

INTRODUCTION

The Web has dramatically changed information seeking and nowhere is this more evident than in the field of medicine. Health information is one of the most frequently requested types of information on the Web. A recent study by Deloitte & Touche and VHA has found that 17.5 million adults in the United States, or 43% of the 40.6 million who use the Internet, are searching for health information [1]. The number of health Web sites is now estimated at 15,000 [2]. Clearly, Americans have both the need and the desire to find answers to their health questions. Many organizations who do not normally provide health information are forming partnerships to add health content to their sites. Health information is now featured at such disparate sites as the *Washington Post*, Netscape, and Cable News Network (CNN).

Although the National Library of Medicine (NLM) has traditionally focused its services on health professionals, the relatively recent explosive growth of the Internet and Web has made NLM resources increasingly attractive to the general public as well. NLM's first Web site, launched in October 1993, described selected NLM programs and provided links to samplers of activities such as the Visible Human Project. In 1996, the site was reorganized and expanded to cover NLM's programs more comprehensively for a wider general audience. Early examinations of the Web search logs showed that users most often searched for names of diseases or other medical terms, probably because they expected the Web site to contain full-text medical information [3].

When NLM made MEDLINE on the Web free of charge in June 1997, the general public immediately became an important NLM user group. Today 30% of MEDLINE searching performed from NLM's Web site is by students and the general public [4]. MEDLINE contains more than nine million citations and abstracts to articles largely written by and for a health professional audience. In early 1998, NLM added twelve consumer health journals to MEDLINE [5] to increase its coverage of information written specifically for the

general public. In October 1998, NLM launched MEDLINE*plus,* a Web-based resource designed to provide consumers with immediate access to authoritative health information. The release of MEDLINE*plus* represented a significant step for NLM in using Web technology to provide consumer-level information. The purpose of this article is to describe MEDLINE*plus*: its goals, development, implementation, supporting technology, current status, and plans for the future.

GOALS

A multitude of Web sites provide consumer health information, but these sites vary greatly in the type, depth, and quality of information they provide [6]. Determining whether the information is reliable and up to date is often difficult. In addition, the Web has spawned duplication of information, with many sites republishing documents from freely available information sources, such as the National Institutes of Health (NIH) and other federal agencies. The increasingly commercial nature of the Web further complicates the search for reliable, unbiased information.

NLM's primary goal in developing MEDLINE*plus* was to help consumers searching the Web find answers to health questions. MEDLINE*plus* was designed to be a well-organized, selective, Web resource that directs consumers to full-text health information available from NLM, NIH, and other reliable sources. In addition, MEDLINE*plus* was designed to assist consumers with searching MEDLINE through preformulated searches on selected health topics. The site's use of consumer language and simple search interface were designed to facilitate use by the general public.

DESIGN AND DEVELOPMENT

As a first step, the MEDLINE*plus* team drafted guidelines for selecting resources. These Selection Guidelines are included as a link on MEDLINE*plus* [7]. NLM chose to link to relevant Web-based information from NLM, NIH, and other government health agencies as a first priority. Additional links to other organizations were selected if the sites included unique information that was primarily educational and not fee-based or commercial in nature.

The prototype design was developed by a small team within the NLM Public Services Division, with assistance from NLM's graphic artist. The focus of MEDLINE*plus* is on national (i.e., United States) resources, not state, local, or international information sources, though there may be exceptions in specialized areas. The sections are:

- **Health Topics**—a list of diseases, conditions, and other medical topics for which NLM has organized links to reliable health information.
- **Dictionaries**—general health-related dictionaries and lists of dictionaries to assist the public with spelling and definitions of medical terms.
- **Databases**—NLM databases and other federal databases that may be useful to the general public. This section also links to the home pages of select Web sites containing comprehensive and reliable information, such as NetWellness, healthfinder, and Mayo Clinic Health Oasis.
- **Organizations**—federal government resources such as the Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC), as well as national associations such as the American Academy of Pediatrics and the American Cancer Society, which have organized information targeted to the general public. Most disease-specific organizations are not included in this section, but are included in the specific Health Topic.
- **Clearinghouses**—federal government and government-sponsored clearinghouses, funded to provide health information to the general public. Other select clearinghouses such as the National Information Center on Deafness at Gallaudet University are also included. Though traditionally these clearinghouses have mailed pamphlets, videos, and other information upon request, they are currently in transition to providing full-text information and databases for direct access through the Web.
- **Publications/News**—federal government newsletters or magazines written mainly for the general public, as well as links to medical atlases, encyclopedias, and major news Web sites with health features such as CNN and MSNBC.
- **Libraries**—libraries offering services to the general public that also have Web sites, including consumer health libraries and selected public libraries; lists of other libraries or library services, such as the National Network of Libraries of Medicine (NN/LM) and U.S. state libraries.
- **Directories**—directories of health professionals and health care facilities.

Each of these eight sections organizes links to reputable Web sites with information useful to the general public. The major focus of MEDLINE*plus,* however, is to identify full-text sources on health topics for reading online, that is, article- or pamphlet-level information. The abundance of such information from the eighteen individual institutes of the NIH provides a solid foundation for MEDLINE*plus.* Other organizations such as the CDC, the FDA, and many national health-related government and professional organizations and associations have been rich sources of excellent information for the public.

DEVELOPING HEALTH TOPICS

As noted earlier, analysis of search terms used on the NLM home page has consistently shown that the ma-

jority of searches (90% or more) were for specific diseases, conditions, or other medical terms. Based on this finding, NLM staff anticipated that the Health Topics section would form the core of MEDLINE*plus.* The original list of health topics was developed from an analysis of a sample of search terms consumers used. Staff of NLM's Lister Hill National Center for Biomedical Communication (LHC) performed mapping and analysis on a five-week sample of searches. The search log contained 87,423 searches that the LHC staff mapped to 56,905 concept terms using the Unified Medical Language System (UMLS), Metathesaurus, SPECIALIST Lexicon, and lexical programs [8]. Of these, over 8,446 occurred more than once and 2,676 could be mapped directly to MeSH terms. For example, shingles, zoster, and herpes zoster were mapped to herpes zoster; herniated disk and slipped disk were mapped to intervertebral disk displacement.

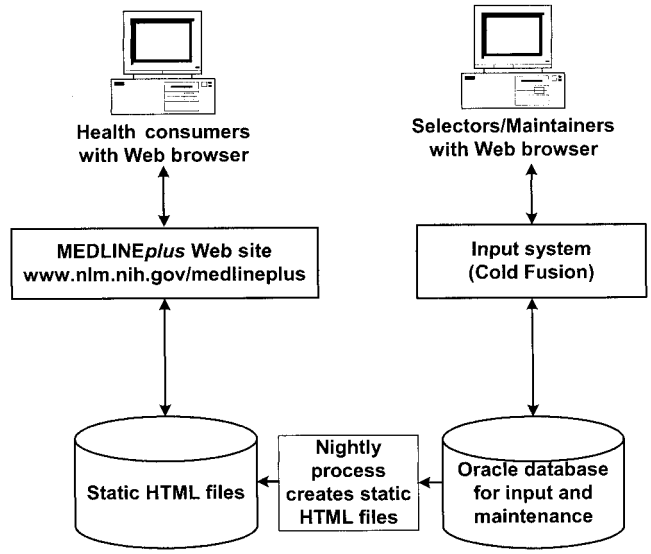
Once mapped, the terms were ranked by search frequency. The most frequently searched topics included diabetes, shingles, prostate, hypertension, asthma, lupus, fibromyalgia, multiple sclerosis, cancer, and other diseases. Drug topics such as Viagra and Zoloft, and alternative medicines such as St. John's Wort were also frequently searched. A target health topic list of more than 300 was developed based on the analysis of user searches to ensure that the health topics of interest to users were included in MEDLINE*plus.*

Each Health Topic page was organized into three major sections: NLM/NIH Resources, including one or more preformulated MEDLINE search; Other Federal Government Resources; and Government and Other Resources by Category. Each Government and Other Resources by Category section may contain up to twenty-one subcategories such as Treatment, Diagnosis, Dictionaries/Glossaries, and Español/Spanish.

NLM's reference librarians created preformulated searches of MEDLINE for each health topic using advanced features of the PubMed search interface. For example, for the diabetes health topic page, preformulated searches on subtopics such as ''reviews, practice guidelines, and clinical trials,'' ''diagnosis,'' ''diabetic foot,'' ''diabetic retinopathy,'' and ''therapy'' were available. These searches when run against PubMed produced a focused list of the most recent citations from English language journals with articles most likely to be readable by consumers. Using PubMed, users could use search techniques such as ''find related'' to locate other articles of interest similar to those presented.

Initially, each Health Topic was constructed as one HTML page. Very early in development, NLM staff realized that if hundreds of Health Topic pages were to be developed, maintaining them as static HTML pages would not be practical. Even with a single topic page, links broke regularly. Within the first eight topics developed, inconsistencies in format and approach

Figure 1
MEDLINE*plus* system architecture



by individual creators became evident. Individual creators pointed to the same resource on different topic pages—for example, heart attack and stroke—but used varying site names and descriptions or pointed to different uniform resource locators (URLs) for the same resource—for example, the home page versus the introduction section to a pamphlet. A more efficient way to build and maintain MEDLINE*plus* was essential if it was to continue to grow and maintain a high level of quality and consistency.

PROCESS AND TECHNOLOGY

The MEDLINE*plus* development team turned to NLM's Office of Computer and Communications Systems (OCCS) for assistance with developing a system for creation, review, and maintenance of MEDLINE*plus.* OCCS recommended using a Web browser compatible database. The database would help ensure consistency of terms and categories and would allow site records to be created and maintained centrally for display on multiple Web pages. The system OCCS implemented used technologies that combined an interactive Oracle database with a Cold Fusion Web forms-based input system. The architecture of the system is outlined in Figure 1.

Using this system, health sciences librarians on contract to NLM contribute to MEDLINE*plus* using a Web browser. The technology allows them to work from any location with Internet access. Currently most MEDLINE*plus*-contributing librarians work in states outside Maryland, where NLM is located. The system architecture has allowed NLM to take advantage of

Figure 2
Top of MEDLINE*plus* entry form: section of form showing site identification and display fields

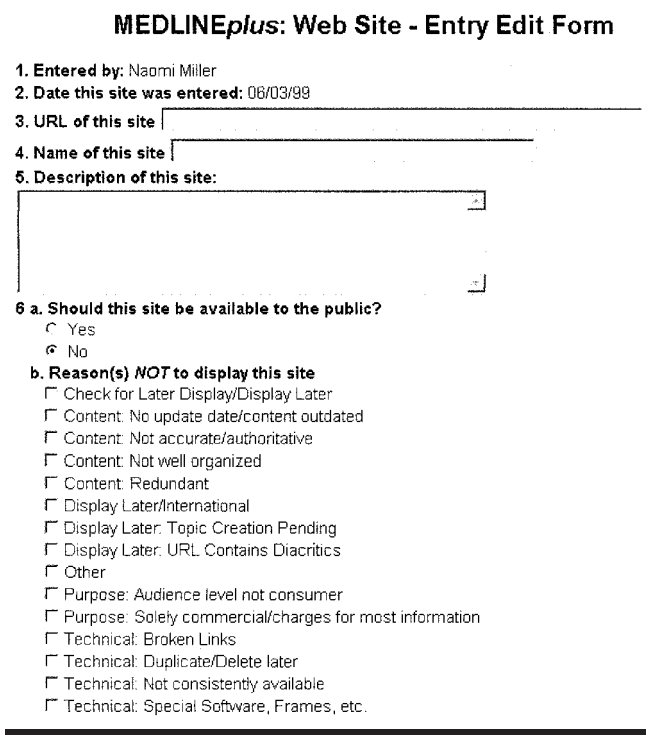
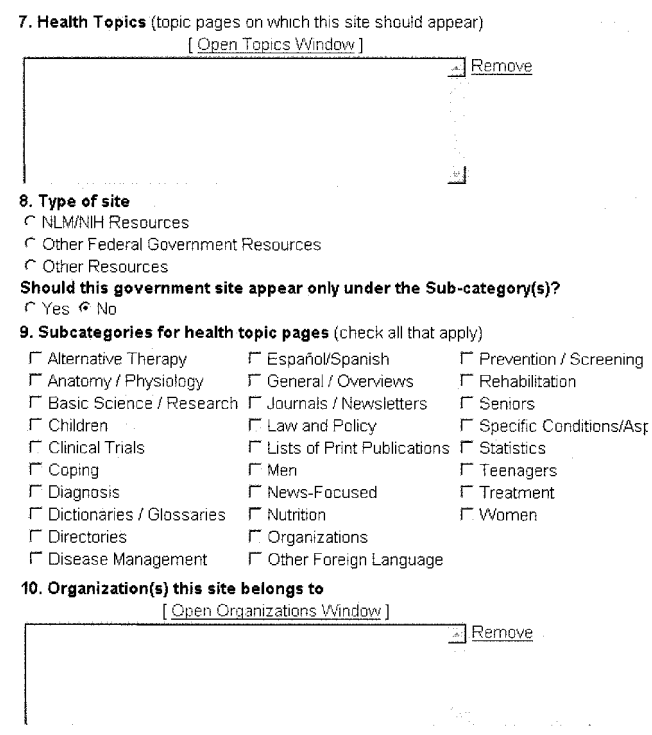


Figure 3
MEDLINE*plus* entry form: section of form showing topics window, type of site, subcategory fields, organization window



expertise at other institutions but maintain the high degree of consistency and quality control afforded by the database.

SELECTION AND REVIEW

The process for creating MEDLINE*plus* records involves selectors and reviewers. Selectors use Web-based forms to contribute records; reviewers approve records for display or return them to selectors for editing. NLM staff and outside contractors are both selectors and reviewers.

Selectors use a Web form to enter records. The form combines free-text fields and fields using lists, radio buttons, and check boxes. Health topics and subcategories are chosen from lists; the more than 300 health topics currently under development can be searched as well as displayed alphabetically. Organization names are subject to authority control. Before the public release of a health topic, a staff member from NLM's cataloging section validates the names of all organizations associated with the records in the database. Selectors may search organization names directly or use a selection list; cross-references are provided to acronyms and variant forms of the organization name. Selectors may enter a new, unapproved organization name, but it will not be released for display on MEDLINE*plus* until it has passed cataloging review.

Successful completion of the input form is dependent upon a set of rules. For example, the system is able to detect a URL that is a duplicate of one already in the database. If a selector has created a record but has omitted a required field (e.g., name of site, type of site, URL), the system prompts the selector to return to the form to complete the required field. An example of a form is shown in Figures 2, 3, and 4.

Reviewers are able to view a list of submitted forms and may view, edit, approve, or return individual records to a selector for reworking. Figure 5 shows a Web form of records pending review. Reviewers may also release topics for public display, move records among topics, manage records for organizations, manage topics by adding topics or changing the names of topics, and manage subcategories by adding subcategories or changing the names of subcategories. The complete administrative menu is shown in Figure 6.

The system used by the development team when reviewing individual records and pages uses the Oracle database. One advantage of this approach is that selectors and reviewers can immediately see the results of approving a site record for display or check on the status of a topic that is still in process. To ensure

Figure 4
MEDLINE*plus* entry form: section of form showing unapproved organization, special site type, and PubMed search fields

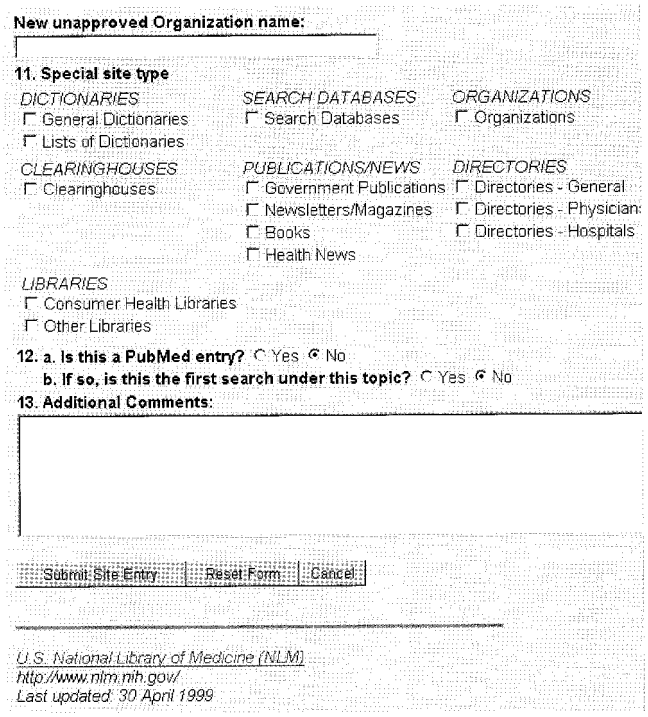
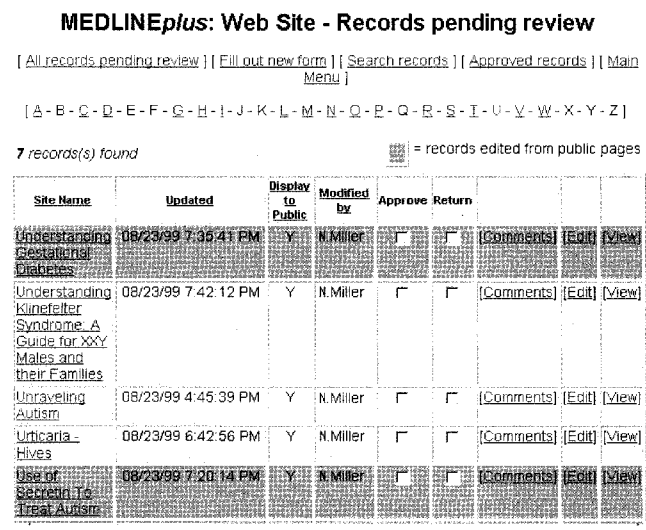


Figure 5
MEDLINE*plus* pending records screen



optimum performance on the public site, however, the public does not attach directly to the Oracle database. The system generates static HTML pages nightly from the Oracle database. These pages are indexed for searching and load quickly on a user's remote browser.

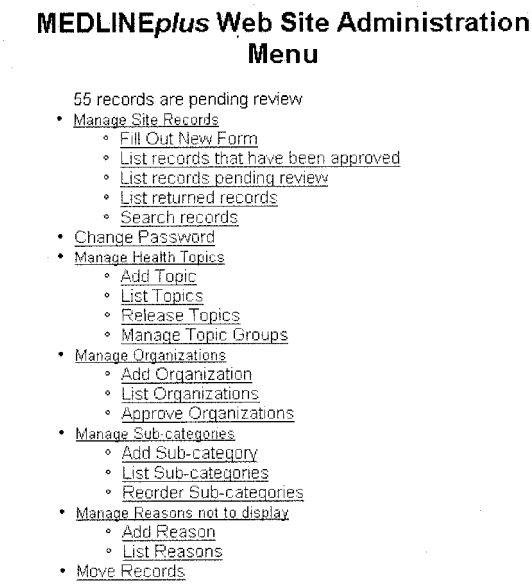
SITE MAINTENANCE

To date, the major effort has been the development of the internal structure and content of MEDLINE*plus*. Through systematic ''combing'' of major government and professional organization Web sites and through the serendipitous discoveries of the many librarians now building this site, the Web site topics have been continuously updated. An automatic link checker is in place to identify broken links, which are fixed daily. Automated Web page checkers that monitor specific URLs for changes to pages are also used to keep MEDLINE*plus* up to date. For example, the NIH ''What's New'' pages are monitored for new documents. Seven months after its introduction, a schedule for systematically updating all sections of MEDLINE*plus* is being implemented. Upon release of a health topic, a release date is associated with the record. These release dates are used to schedule periodic updating. All major sections of the site are continuously updated, but a complete review will be done every six months.

CURRENT STATUS

MEDLINE*plus* made its official debut on October 22, 1998, in conjunction with NLM's consumer health pi-

Figure 6
MEDLINE*plus* site administration menu



lot project with public libraries. The pilot project was NLM's first major initiative to increase the public's awareness of and access to health information on the Internet. NLM has worked with thirty-nine public library organizations at some 200 sites in five states and the District of Columbia. NLM, in partnership with medical libraries, has provided training and support to librarians in searching MEDLINE and MEDLINE*plus* so that they can better assist patrons in finding health information.

NLM has obtained feedback on MEDLINE*plus* from a number of sources. Feedback from the public libraries participating in NLM's pilot project has been positive. NLM has implemented changes based on suggestions from users, such as shortening the Health Topics pages by listing some government Web sites only under the categorized list. The librarians and patrons who have used MEDLINE*plus* have consistently reported that as more topics are added, the site becomes more useful to them. At this writing, the health topics covered have increased from twenty-two to more than eighty, and this number is expected to grow quickly to complete the identified topics. A ''feedback'' button has recently been added to the site and dozens of topic suggestions have been received, along with numerous compliments and other suggestions.

CHALLENGES

Experience with and feedback on MEDLINE*plus* have supported NLM's plan to continue its development and to begin to promote this service nationwide. Though users are generally very pleased with the organization and authoritative information provided through MEDLINE*plus,* there is room for improvement. The most common recommendation from users is that MEDLINE*plus* should contain more topics. Completing over 300 health topics with preformulated MEDLINE searches by the end of 1999 is a priority.

Complicating development is the paucity of authoritative Web information on some of the topics most frequently searched by consumers. At the same time, avoiding duplication is difficult when multiple Web sites post the same information. Sometimes different Web servers from the same organization will publish the identical document with different URLs. Though NLM does not plan to create new information content, NLM does plan to encourage NIH and other appropriate health organizations to create information for access through MEDLINE*plus.*

NLM has continued to analyze the terms input by users of the NLM main Web site and the MEDLINE*plus* site to keep up on what users want. The MEDLINE*plus* search engine, ht://Dig [9], offers ''fuzzy'' searching and soundex options. A search for ''Alzimers'' presents the user with the link to the Alzheimer's Health Topic page. NLM staff are optimizing the ht://Dig search engine and testing other search engines to optimize retrieval. An analysis of the terms has provided the development team with an insight into the difficulty many Web users have with the medical language. In a follow-up study of over 300,000 queries of the NLM Web site, NLM's Lister Hill Center staff has reported numerous examples of misspelled words. Words such as fibromyalgia, gynecology, and prostate are often misspelled. Eponymic terms such as Crohn's disease are misspelled in a number of ways. NLM will explore the development of a terminology server to mediate between user terminology and the terminology used in medical sources [10].

NLM has begun to investigate whether the database methodology being developed for MEDLINE*plus* can be used by other organizations organizing health information on the Web for other audiences. Shared records, vocabulary, and other tools, as well as authority work done by NLM, may prove feasible to reduce redundancy among efforts and allow other organizations to combine what NLM has done on a national level with consumer health resources at the local level.

The major development for MEDLINE*plus* will be completed during 1999, but as with the Web itself, NLM expects MEDLINE*plus* to continue to expand as a dynamic, useful resource for the general public, health sciences librarians, and health professionals.

REFERENCES

1. DELOITTE & TOUCHE, VHA, INC. 43% of consumers use Internet for health information. [Web document]. Irving, TX: Deloitte & Touche, April 5, 1999. [rev 8 Apr 1998; cited 7 Sep 1999]. <http://us.deloitte.com/us/news/99april/eastudy.htm>.
2. GROVE AS. Internet health day keynote speech. [Web document]. Intel Corporation, 1999. [rev 1999; cited 7 Sep 1999]. <http://www.intel.com/pressroom/archive/speeches/ag102798.htm>.
3. MILLER N. Improving the NLM home page: from logs to links. Poster presented at Medical Library Association annual meeting, May 26–27, 1997. [Web document]. Bethesda, MD: National Library of Medicine, 1997. <http://www.nlm.nih.gov/psd/web_poster/web_poster.html>.
4. Online usage statistics smashed; free MEDLINE rewrites NLM record book. NLM Newsline 1998 Jan–Mar;53(1):1–2.
5. MEDLINE to index selected consumer health publications. NLM Newsline 1998 Jan–Mar;53(1):2.
6. JADAD AR, GAGLIARDI A. Rating health information on the Internet: navigating to knowledge or to Babel? JAMA 1998 Feb 25;279(8):611–4.
7. NATIONAL LIBRARY OF MEDICINE. MEDLINE*plus*: selection guidelines. [Web document]. Bethesda, MD: The Library, 1998. [rev 22 Jan 1999; cited 7 Sep 1999]. <http://www.nlm.nih.gov/medlineplus/criteria.html>.
8. NATIONAL LIBRARY OF MEDICINE. Unified Medical Lan-

guage System. [Web document]. Bethesda, MD: The Library. <http://www.nlm.nih.gov/pubs/factsheets/umls.html>
9. HT://DIG. WWW search engine software. [Web document]. The ht//Dig Group, 1995–99. [rev 1 Jun 1999; cited 7 Sep 1999]. <http://www.htdig.org/>.
10. MCCRAY AT, LOANE RF, BROWNE AC, BANGALORE AK. Terminology issues in user access to Web-based medical information. AMIA 1999 Annual Symposium, Washington DC, November 8, 1999, Session S-16.

Received June 1999; accepted August 1999