INTRODUCTION

This paper discusses the design, development, and evaluation of GRATEFUL MED, the National Library of Medicine's (NLM) front end software for microcomputers that was developed to assist physicians and other health professionals to search NLM's MEDLINE database. A search is constructed by filling out a form screen with information on the desired author, title, and/or subject(s); the search can be limited to English language, review articles, or a particular journal. No knowledge of Boolean connectors or the Library's Medical Subject Headings (MeSH) vocabulary is assumed. The search is constructed and the results reviewed on the user's microcomputer; that is, while not connected to the NLM mainframe.

By March, 1988, 9,500 copies of GRATEFUL MED had been sold. The owners comprised 38 percent of the MEDLINE searchers (by codes in use), accounting for 30 percent of NLM's mainframe computer usage.

Throughout the relatively short life of GRATEFUL MED, its designers have worked toward the principle of responding to the users' expectations while adhering to the original design philosophy—not always an easy task. This chronological look at the development of GRATEFUL MED shares some of the practical "lessons" learned along the way that may assist others interested in the design and evaluation
of the computer/human interface. These lessons are presented from the personal perspective of the first author, not NLM.

"Design philosophy," in this context, refers to software functionalities rather than the ADP definition of what takes place between analysis and implementation. Reconciling the design philosophy and user expectations thus might be better phrased as "adjusting what you have decided to provide with what functions users (grow to) expect." It sounds rather recalcitrant when phrased this way; however, if the design philosophy is abandoned as a result of responding to the often diverse user recommendations, the software develops in a hodgepodge way, resulting in a deviation from the original goals. The key is "reconciling"—answering the user's need within the framework of the design philosophy.

THE BEGINNING

The first author's daily notes for December 14, 1984 say: "Upbeat meeting with John Anderson [Director of Information Systems] and Lois Ann Colaianni [Associate Director of Library Operations], and others. Dr. L. wants a concrete proposal for end user services."

Dr. L. refers to Donald A. B. Lindberg, M.D., Director at the National Library of Medicine. This was one of his first directives; he had been sworn in only two months before. The author's specific assignment from this meeting was to survey existing end user search systems, including mainframe resident, microcomputer front ends, and front ends co-resident on a mainframe—both operational and prototypes—and prepare a report on the state-of-the-art that included recommendations for directions NLM might take in end user searching. The report was to identify existing commercial systems for possible endorsement or purchase by NLM.

Dr. Lindberg emphasized that the NLM system goal was to develop a searching mechanism for the end user that would obviate the need for the user to have experience with the three stumbling blocks to searching: logging on, using the ELHILL command language, and using the NLM controlled vocabulary, MeSH. In addition, it would be easy to use, inexpensive, and allow for a growing searching sophistication if the user so desired. He expressed a preference for a system that would require minimal documentation, perhaps a few pages, with extensive online help. Plans were for the evolutionary development of successive versions of the software, each "smarter" than the preceding one.

By February 2, 1985, the author's office walls were plastered with huge charts describing twenty-four operational and prototype systems,
each described by a number of criteria, such as mode (command or menu), search logic, and key system features and capabilities. One prototype, called MICROSEARCH, was especially interesting, for it was designed to interface with software that had been derived from the NLM mainframe search software, ELHILL, which is used to search the NLM family of databases. Importantly, it appeared to satisfy all of the desirable characteristics identified by Dr. Lindberg—and then some.

All of the intellectual processes were done while the user was sitting at his microcomputer but not connected to the host mainframe computer. This feature was appealing because it freed the user from the sense of urgency that often occurs when the "clock is ticking" and charges are mounting. Automatic logon, built-in Boolean connectors, and an algorithm to recommend subject headings based on the user's judgement of the retrieval were other features of import. The design also included the concept of extensive online help. At the conclusion of the study in mid-February 1985, it was recommended and agreed that NLM would undertake with Online Information International the adaptation of the MICROSEARCH prototype to the NLM environment, a not-especially overwhelming task given its common NLM/ELHILL origin.

ACCEPTANCE TESTING OF SELECTED PROTOTYPE

The Office of Planning & Evaluation (OPE) was enlisted with providing a process to gather end user input for the evolving set of requirements for a MEDLINE front end based on the prototype. But before that, the NLM team had to do in-house acceptance testing of the MICROSEARCH prototype as it took on the look and feel of what was to become GRATEFUL MED. A highly interactive team approach between NLM and the prototype developer was thus begun at this early stage and continues to this day. This has been a major factor in the rapidity of the development. Later in the process, Online Information International was awarded a competitive contract to continue development, and is referred to from now on as the contractor.

Acceptance testing of the prototype was done by a small in-house group from MEDLARS Management Section (MMS) and OPE over an eight-week period. The group went through twelve versions in this short time, and major changes were made—cosmetic, functional, and philosophical. Written reports were compiled from testers' comments, and the contracter responded with feedback and fixes as rapidly as problems were found. Proposed enhancements were discussed by telephone and electronic mail. Thirty-two problems were identified and corrected; nineteen enhancements were considered, many of which were implemented and found to be useful.
Adherence to the design philosophy was a struggle throughout, even at this early stage when feedback was limited exclusively to in-house staff. For example, the team strongly considered allowing the user to limit a search to HUMAN, a frequently used search tactic in clinical searches (to eliminate animal research). This can be done by combining HUMAN, a legitimate MeSH term, with a search by a Boolean AND. It can also be done by combining AND NOT ANIMAL with a search. The front end could be constructed to do the search either way. However, experimentation showed that the more straightforward AND HUMAN increased the cost of a search sharply due to the ANDing of such a highly posted term. Searching with AND NOT ANIMAL, on the other hand, lost from 0 percent to 25 percent of the retrieval in a group of test searches—those citations indexed to both HUMAN and ANIMAL. A totally unexpected finding was that otherwise relevant citations indexed to neither HUMAN nor ANIMAL included a large number of articles (33 percent of the total) classified as "review articles." Since the software was intended for end users, many of whom would be doing searches to retrieve "a few good articles," it was felt that this potential elimination of review articles was likely to prove harmful. It was also felt that searches of a clinical nature frequently are limited to HUMAN by the very nature of the other terms in the search, making it unnecessary to add the term. An added factor was the significant workload that would be added to the mainframe if large numbers in the health professions adopted the front end for searching and elected to routinely (and simultaneously) limit their searches to HUMAN. Ultimately (this whole process spanned perhaps two days), it was decided to provide the user with an easy option to limit the search to REVIEW rather than HUMAN.

This rather lengthy example is given to illustrate one of the most important things learned:

*Lesson #1: Test every idea for enhancement thoroughly, using a group of searches if necessary, to measure the effects of the change—especially the unexpected effects.*

Each enhancement or change was thoroughly considered against the design philosophy before implementation in the evolving front end. This was to become the pattern throughout development of the ensuing versions.

Dr. Lindberg and other senior staff were shown the prototype several times, both routinely and when the group was dealing with difficult decisions. In addition, the prototype was demonstrated to several members of the NLM Long-Range Planning Panels (both health professionals and librarians), the NLM Board of Regents, and selected visitors
of NLM senior staff. Some of these individuals were recruited as on-the-spot testers, and their reactions were invaluable. The acceptance testers had already grown too familiar with the prototype and were missing some obvious places where there was room for improvement. This is a natural tendency as patterns of use develop, which leads to:

Lesson #2: Add new testers periodically at all phases of development to avoid the problem of "pattern testing."

A recommendation for acceptance of the prototype was made in early October 1985. The front end was tentatively named "MEDSEARCH," and all on-screen and written references as well as the draft documentation used this name.

PROTOTYPE TO VERSION 1

Formal Beta Testing

A MEDSEARCH Beta Test Group was appointed by John Anderson in late September as acceptance testing was coming to a close. Besides the authors, Becky Lyon-Hartmann, Regional Medical Library Coordinator, and Edward Sciullo, from the Office of Computer & Communications Systems, rounded out the team that was to develop and carry out a Beta test plan.

More explicitly articulated design principles for both the software and the Beta test were needed to provide to the Beta testers, and ultimately, to the prospective users. They were produced as follows:

"The goal of MEDSEARCH is to provide a microcomputer software tool for end users and librarians that will:

— enable MEDLINE and CATLINE searches to be done with little or no knowledge of the presently required login procedures and search mechanics, or of the controlled vocabulary (MeSH);

— produce a reasonable number of citations that are responsive to the user's information needs and offer guidance in a more comprehensive search, if so desired;

— upload a search and download the results to minimize online time and attendant charges; and

— encourage, but not require, the user to learn more sophisticated searching techniques."

"The goals of the test were, in order: to pre-test the current
prototype software to discover serious flaws or technical deficiencies prior to widespread distribution to the biomedical community; and, to obtain suggestions for future improvements and enhancements."

The Beta Test Group elected to use a phased approach to the Beta test. Phase I would consist of in-depth structured interviews and observations of actual search sessions with a cross-section of end users and librarians at each of three local test sites. In Phase II, the software would be distributed to up to fifteen sites for wider, but less tightly monitored, testing. The types of information to be obtained, in whole or in part, at each of the Beta test sites included:

— who the users are (professional role/specialty; computer experience);

— who the non-users are, and why;

— location of system (library reading room, hospital, laboratory, home, office, etc.);

— types of searches performed;

— user satisfaction with information retrieved;

— user satisfaction with the process;

— use of follow-on search capability;

— use of advanced front end and native search capability;

— problems, if any, in establishing log in;

— inadequate, incorrect, or inefficient documentation; and

— problems, if any, in installing software.

For the sake of consistency, one person was appointed to conduct the Phase I test using a packet prepared for collecting demographic information about the user, detailed information on the search itself, observations of the user/system interaction, information on user satisfaction, and suggestions and comments from the user. Structured questions were developed for the demographic and user satisfaction information; forms were outlined for the remaining open-ended information
to be recorded by the observer. A two-minute "script" was prepared in order to deliver the same introduction to each test subject.

The NLM Reading Room was used as the initial site for Phase I Beta testing in order to pre-test the questionnaire and procedures. This goal was met, and a few minor changes were made to the questionnaire; however, the user population volunteering for the test at this site was composed of a disproportionate number of students in non-health fields and "private citizens," neither of which was the target health care professional population for which the system was conceived and designed. The reaction of the intended user population was anxiously awaited, and led to the rather obvious, perhaps,

Lesson #3: Test with the intended user population for the most useful feedback.

The two remaining Phase I sites, the University of Maryland Health Sciences Library and Fairfax (Virginia) Hospital Medical Library, coordinated by Frieda Weise and Alice Sheridan, respectively, provided useful feedback that confirmed that it was appropriate to proceed with Phase II. Thirty-five subjects at these two sites performed searches of their choosing without significant problems. The Beta test questionnaire that was to be used at the Phase II remote sites was also pre-tested with these users. There were a number of suggestions for enhancement, all duly recorded. It had been speculated that the information needs of the end users at the research site (Maryland) and the clinical site (Fairfax) would be quite different, and this was borne out. The clinical subjects were less interested in extensive retrieval (a "few good citations" sufficed), more time conscious, and keen on the idea of searching at home in the evening after seeing patients.

More feedback was wanted before deciding what substantive changes to make to the software, but there was enough observed behavior to make six small changes. One change did not seem particularly significant at the time—the addition of a Help screen for a search that retrieved nothing. This problem of zero retrieval became the focus of much more attention later on, thus:

Lesson #4: Pay close attention to every problem the user has; in the early stages, it is difficult to recognize what is really important.

Examples were changed and added in the User's Guide, which was truly an example of minimalism. Some of the expanded examples were searches done by Phase I testers; it added a note of realism to know that someone actually wanted information on these subjects.

Phase I testing was finished on November 25, and Phase II testing began. The seven Regional Medical Libraries (RMLs) had been asked and had agreed to participate: New York Academy of Medicine; Uni-
versity of Maryland Health Sciences Library; University of Illinois at Chicago, Library of the Health Sciences; University of Nebraska Medical Center, McGoogan Library of Medicine; University of Texas Health Science Center at Dallas Library; University of Washington, Health Sciences Library; and UCLA Center for Health Sciences, Louise Darling Biomedical Library. In addition, the following institutional sites had volunteered to participate in Phase II: McMaster University, Department of Clinical Epidemiology and Biostatistics, and Health Sciences Library; University of Missouri, Information Science Group, School of Medicine; University of Pennsylvania, Biomedical Library; and University of California, San Diego, Biomedical Library. In addition, several individual end users had volunteered. Some of the latter group would be sharing the software with others at their sites, including one clinic setting with sixteen staff and forty residents. Each site was contacted by telephone for a discussion of the goals, test methodologies, and timeframe. By November 27, the Beta test group was formed and the test was scheduled for mid-December through January, 1986.

At this point, the official name for the software was selected by Dr. Lindberg from a list of twenty-two staff suggestions. (MEDSEARCH, having been registered by a firm that had produced other front end software, was not eligible.) The new name, GRATEFUL MED, produced mixed reactions from the day it was selected, but it has proven to be a happy choice, provoking much interest and easy recall. The Beta test materials were quickly modified to reflect the new name.

The library/clinic system test sites were given the option of observing twenty-thirty uses of the software or leaving a self-administered questionnaire for anyone using the system to conduct a search, or a combination of both options. The site coordinators were also requested to provide a one or two page summary of their impressions of the system and its use at their institution. The individual end-user volunteers were asked to either complete the questionnaire or submit written comments summarizing their experience. Free access was provided for all searching and a “help” telephone number provided.

Except for the demographic information, the questionnaire used in Phase I contained open-ended questions that were completed by the observer. To increase the ease with which the self-administered questionnaire could be completed, it was restructured and reformatted to contain mostly multiple choice questions; open-ended questions (comments and suggestions) were optional unnumbered items on the last page of the four-page questionnaire. The draft documentation and copies of the questionnaire were sent out with the software on a floppy disk on December 13, 1985.
Beta Test Results

By February 1986, reports of over 600 documented uses of GRA\-TEFUL MED had been received from ten of the Beta test sites. Coincidently with the development of the front end software, the National Library of Medicine was celebrating its Sesquicentennial Year. The first major event was a symposium for 100 medical writers and journalists scheduled for February 5, 1986. One of the goals of the symposium was to acquaint the attendees with the services of the NLM that might assist them in their writing, and it was suggested that a Beta test version be distributed at the symposium. External time pressures were now a factor. First, if major problems had been discovered by the Beta testers, it was important to reverse the decision to distribute at the Science Writers’ Symposium; second, if distributed as planned to the writers and mentioned by them in print as forthcoming, the release of Version 1 should follow closely.

The Office of Planning & Evaluation undertook the task of data analysis, both from the questionnaires and the site coordinators’ summary reports. From the standpoint of stability and robustness, GRA\-TEFUL MED stood the test. From the standpoint of enhancement, the Beta testers had provided even more data than originally expected.

There was a wealth of valuable suggested improvements that would enhance future versions; more importantly, however, a consistent pattern of frequently requested changes clearly emerged. Up until now, the changes and enhancements to the evolving front end had been suggested primarily by the contractor and NLM staff—chiefly the small number of NLM staff involved in acceptance testing. This was the first set of recommended changes coming directly from a large group of potential users, and the process of reconciling the design philosophy and user expectations began in earnest. Some of the requested changes fit nicely with the team’s goals—some did not. It soon became apparent that the front end could not serve all populations equally well. The librarians’ requests reflected their in-depth knowledge of searching. The end users’ requests reflected their lack of knowledge. It clearly would be impossible to “be all things to all people,” and imprudent to try.

It also became apparent that all portions of the design philosophy had to be flexible. For example, the principle of uploading the search and downloading the retrieval, and keeping the intellectual processes of the user tied to the microcomputer, not the NLM mainframe, was questioned by one library site in particular. The librarian wanted the option to interrupt the search during the uploading or downloading for the purpose of modification, an act that required knowledge of and experience with the NLM command language. A direct logon to the
NLM computer was allowed through the initial menu, and it was assumed the librarian users who might be making use of the front end as an auto-dialer would elect to use that option. Consideration had not been given to the idea that librarians might want to take advantage of the user-assisted form screen with the option to later modify. Although it was contradictory to the design principle, this capability was added. It was felt that the risk to the end user was small—the occasional end user who might accidently invoke this feature would receive a message including instructions on how to send the “stop command.” It was also possible that an ambitious end user might absorb enough from observing the uploading and downloading actually to make use of this real-time modification option. This fit with the design goal of providing the user the ability to “grow” with the system if he or she so desired.

Lesson #5: Even the most sacred design principle can be flexible.

Another example that compromised a point of philosophy involved the display, consideration, and printing of each citation in a retrieved set. The algorithm in GRATEFUL MED that was to guide the end user to appropriate MeSH terms for an expanded search on the same topic was not nearly as appealing to the end user in a hurry as it was to the designers. The student testers, especially, wanted to “print and run” rather than give serious consideration to each citation and print only those deemed relevant to the search question. (One physician suggested that it was not so much the time, but rather that the students did not yet have the experience to know what was relevant.) Of those who were willing to go through the retrieval one by one for printing, many did not want to see or print the MeSH terms. Although a goal of the design philosophy was, and is, to provide MeSH assistance, options were provided to skip viewing and printing MeSH terms, and indeed, to skip the review process altogether. This last change was implemented in the only way possible at such a late date—as a choice in the relevancy question. Because the screen space is limited, the instruction is terse, and perhaps not the optimum implementation.

Lesson #6: Regardless of how much one wants to be responsive to users’ expectations, one must weigh carefully whether it is worth making a last-minute change; it might be better to save the change for a future version and implement it differently.

Eight recommended changes warranted serious consideration prior to release of Version 1. The contractor was able to incorporate six of them very quickly. One was already planned for Version 2 (providing MeSH headings during the construction of the search query—a major undertaking that added a second floppy disk to the software). The last recommendation was to put portions of the help screen for zero retrieval
(newly added with Phase I testing) directly into the "zero hits" message received for such a search. That is, do not require the user to ask for the help screen, but provide it without his asking. Since the message would get stale with enforced repetition, this suggestion was not implemented, and a better solution to this problem is still being sought.

Three other changes were added in the category of "niceties" rather than "imperatives." It was necessary to issue a sheet titled NEW FEATURES OF GRATEFUL MED, VERSION 1.0 with the User's Guide, which had already been sent for printing. This was a problem, as several users did not realize the significance or importance of this two-page addendum. In retrospect, a better approach would have been to print it on bright colored paper and title it BIG, NEW, IMPORTANT FEATURES THAT ARE NOT IN THE USER'S GUIDE!!!

Lesson #7: Do not assume that the users will read every piece of documentation included in the packet. If it is important, make it appear so.

GRATEFUL MED, Version 1, was officially available for order from the National Technical Information Service in March 1986 for $29.95. It had taken slightly more than a year to bring the front end from prototype to release.

Feedback from Library Users

An NLM Fact Sheet with the header "GRATEFUL MED: A New Way to Search MEDLINE" was produced and widely distributed. The secondary header, "System Intended for the Individual User," indicated the proposed user population. This proposed user was originally visualized as the individual physician or other health care professional, at home, at night, with his or her microcomputer, doing searches for the patients seen that day. The clinicians among the Phase I Beta testers had reinforced this thinking by their comments. It was, therefore, a happy surprise when a number of medical libraries bought copies and integrated end user searching into the public services/reference setting, with the cost of searching paid by the library budget.

End user instruction for GRATEFUL MED was a natural outgrowth at some of these institutions, and feedback from the librarians and health professionals who have written abbreviated instructions and "brief guides" has been invaluable. Data from the Phase II Beta test showed that, generally speaking, not much attention is paid to written documentation. Material was added to the online help screens where the need for additional instruction was indicated, expanding the Version 1 User's Guide only slightly from the one used in Beta testing. The instructional materials developed by library users served to show what some users thought was needed in addition. Better documentation was
needed for those who would use it; thus Chris Olson & Associates, an Annapolis, Maryland firm experienced in the production of end user materials, was hired to create a logo and produce both a brochure and a "slick software manual."

Lesson #8: Accept the fact that end user documentation is difficult to write and rarely read. Then try to produce the best possible for those who do read or refer to it, making use of instructions written in the "field" by actual persons using the software.

Feedback from Individual Users

About 25 percent of the requests for new access codes to the NLM mainframe in the first few months of availability of the front end were from individual end users who had purchased GRATEFUL MED (this had grown to 36 percent by August, 1986). Suggestions, some in lengthy letters, were received from a number of end users; they were added to the already growing list of possible enhancements. It became obvious quickly that there were some sophisticated, computer literate end users who wanted the front end to stretch and grow. There were also naïve users who had no real interest in learning more than was needed to do a simple search in Version 1. Both groups were important to satisfy.

Lesson #9: If at all possible, do not force users to learn new or changed features in order to use the software the way they first used it. But make the new and changed features appealing so they will want to try them.

VERSION 1 TO VERSION 2

As is often the case, work began on Version 2 before the release of Version 1. Online Information International, Inc., in the continuing role as contractor to NLM, delivered a test copy of Version 2 to NLM three months after the release of Version 1. The two major thrusts of Version 2 were the addition of a permuted MeSH list from which search terms could be selected and the capability to download new versions and/or features from a commercial mainframe (the Source). Both were achieved, but ironically, having the first precludes much of the use of the second.

The ability to select from MeSH had been requested by almost everyone who used the software who had previously used either the printed Index Medicus or the NLM's command language online search system. NLM's thesaurus for both indexing and cataloging, MeSH is a highly developed, hierarchical list of over 14,000 terms. These are further divided into major descriptors (appear in Index Medicus), minor descriptors (do not appear in Index Medicus), and entry terms (cross references to major descriptors that appear in the printed MeSH, but
not in *Index Medicus*). The goal was to fit as many of these terms, in a permuted fashion, onto one double-density, double-sided floppy diskette—362,496 characters. Even with compression techniques, it was impossible to include the entire range of the terminology.

To now deviate from using MeSH in its complete format was disconcerting to some staff, but it had to be done. Other than agreement that major descriptors could not be eliminated, there was diverse opinion on how to handle this problem. Alternatives for omitting all or portions of the minor descriptors and cross references were presented and, after discussion, the team opted to exclude the chemical terms from the minor descriptor and cross reference lists. This allowed for the inclusion of minor descriptors and cross references from the remainder of MeSH. The decision was based largely on the needs of the clinician end user community, for whom the extra chemical terms were perceived to be of less potential use than other categories of terms (for example, the disease terms).

*A Lesson #10:* Remember the population to be served, and make design decisions based on best serving that population.

The principle of downloading new modules or versions, which was implemented in Version 2, is still in favor. MeSH, however, is also included on a second diskette. As a result, new versions have been released by mail in conjunction with the annual update of the MeSH vocabulary and the production of a new manual, since it quickly became apparent that it was counterproductive to use the download feature for the dissemination of two full diskettes. The time (to the end user) and the cost (to NLM) are prohibitive. A much better use is for updating or correcting a smaller program module. This has been done and the expectation is that the feature will be used more in the future for this purpose.

*A Lesson #11:* When two major, sweeping enhancements are being tested, it is important to test both simultaneously.

Altogether, nineteen of the twenty-eight suggestions changes on an ever-growing list of potential enhancements/changes were implemented.

**More Beta Testing**

There was no shortage of volunteers for Beta testing Version 2. Goals for this test were to discover problems in the software and assess the usefulness of the new features. It was decided to retain a few of the Version 1 testers for continuity and enlist some entirely new ones. While testing Version 1, it was learned that it was important to include both institutional and individual users in a Beta test. Individuals can
provide excellent insight and well thought-out suggestions for enhancement, but their overall volume of searching is not sufficient for uncovering "bugs." Institutional testers can provide both suggestions and volume use, both dependent, of course, on the individual coordinating the test.

*Lesson #12: Include a variety of users in Beta test groups in order to obtain an adequate amount of testing and users who will do insightful searches.*

The final constitution of the Beta test group included the following institutional sites as repeats from the Version 1 group: McMaster University, and one of the Regional Medical Libraries, UCLA Center for Health Sciences. New institutions included: the Uniformed Services University of the Health Sciences; Cedars-Sinai Medical Center, Los Angeles; Indiana University School of Medicine; Catholic University School of Library & Information Science; and Ohio State University, Microcomputer Laboratory. The Beta test group also included individual users, some of whom shared the software with others at their workplaces. They received Version 2 in late October, 1986 and tested through December.

In-house staff from MMS had spent about thirty hours in organized testing before Version 2 was distributed, and only one additional "bug" was uncovered by the Beta testers. There were three other important consequences of this Beta test. First, it was discovered that "IBM-compatible" did not always hold true; a number of problems with IBM clone machines were discovered by the testers. Secondly, the much-labored-over decision involving the makeup of MeSH in GRATEFUL MED was confirmed as correct. That alone was worth the test. And third, at the urging of some testers and Version 1 users, two features were added, even though the documentation had already been sent to the printers. One feature was at the request of librarians, and the other, for the health professional users.

The first change was to provide a way to suppress the downloading of MeSH headings (the librarians' request) to reduce the cost of searching. Experiments revealed that the total cost of a GRATEFUL MED search could be reduced 20-50 percent by the omission of MeSH from the downloaded retrieval. This was due to the charging algorithm used at NLM. Suppression capability could be provided for librarians subsidizing end user searching in libraries, but how to do so posed a problem. There was concern that, given the opportunity by a direct question, novice users might elect to "skip" MeSH terms without understanding what they were. The decision was made to implement the function in a subtle rather than overt way, resulting in perhaps the most criticized enhancement undertaken to date. Most users do not
know the suppression option exists; of those that do, some do not like the implementation.

Lesson #13: If it is necessary to "protect" the users from a capability, further consideration should be made as to the validity of the change as well as methods of implementation.

The second late addition was for the end users: a way to write citations to a named PC disk file, either new or previously created. This was much appreciated, easily understood, and popular with end users and librarians alike.

Lesson #14: Make a last-minute change only if it is well thought out and intuitively understood without documentation.

Institutionalizing the Product

One of the normal evolutionary changes that accompanies any effort of this type began during Version 2: the institutionalization of the product. What started as a few people in a new venture became a structured group with explicit procedures. This evolved by three steps over a six-month period.

The Office of Computer and Communications Systems (OCCS) had provided management and funding support since the beginning of the project. The Director of OCCS had personally coordinated the development. The first step in building a structured group was in April 1986, when the Director phased himself out and appointed Philip Nielsen as Project Manager and Chair of the Working Group for GRATEFUL MED.

In May, the Associate Directors from the three program areas at NLM involved in user services instituted a procedure to prioritize the "wish list" for enhancements. The Associate Directors from Library Operations and Specialized Information Services join the Director of OCCS in regular meetings to consider the more formally named "GRATEFUL MED Priority List." Convening this group provides an opportunity for management to weigh the costs and benefits of the proposed changes and assign priorities before the list is submitted to the Director of NLM.

The third step in the institutionalization came in October 1986, when an official GRATEFUL MED Working Group was established, composed of representatives of several Library areas, and the contractor, whose suggestion it had been to form a Working Group.

This Working Group since that time meets biweekly, often for three hours or more. In terms of staff time, it is an expensive venture. More diverse perspectives lead to lengthier decision-making processes. Overall,
however, these meetings are a workable method for maintaining a team approach.

Lesson #15: To encourage staff participation in and acceptance of a new product, involve all program areas in the institution who have a vested interest in the outcome, both at management and working levels.

Integration into Operations

GRATEFUL MED was further institutionalized when it was installed in the NLM Reading Room for online access to the card catalog (CATLINE) as well as for MEDLINE (journal) searches. Since NLM library patrons, many of whom are repeat users, were accustomed to the natural language CITE system for CATLINE, a gradual transition was provided to GRATEFUL MED. The outdated hardware used for CITE was impossible to replace and increasingly difficult to have repaired. One microcomputer was installed in the Reading Room and user aids were developed for both the software and the PC. A second PC was installed several months later; in November 1987, all of the CITE terminals were replaced with GRATEFUL MED on PCs.

Staff of the MEDLARS Management Section were also heavily affected by the software, particularly by increases in the number of inquiries for information, paperwork involved in adding new users to the system, and service calls for assistance. There has been a marked increase in all of these activities, especially immediately following a new release. In addition, with Version 2, MMS was named directly responsible for in-house testing, a time-consuming task.

Reference and MEDLARS Management are two groups in the Library who observe or talk with end users directly about the software on a regular basis. Both have representatives on the Working Group, facilitating the transfer of direct observation and experience to the evolution of the software.

Lesson #16: Provide a means for staff who regularly interact with the users to routinely report feedback.

One huge facet of software production is documentation. The original idea of providing only minimal paper documentation and maximum Help screen information was finally modified by users who repeatedly requested a “real manual.” The documentation contractor had designed a total software package, including a three-ring User’s Guide with diskettes in vinyl pockets, all in a slipcase. A subset of the Working Group had worked closely with the contractor to produce the text; layout and formatting were done by the contractor. The project remained on schedule until a change in the regulations at the Department
of Commerce, the parent organization of the National Technical Information Service, caused a three-month delay in printing.

*Lesson #17: Allow extra time in the schedule for all portions of the production that are not in your complete control.*

Version 2 was released in March 1987, thirteen months after Version 1.

**VERSION 3**

**GRATEFUL MED,** by now, had become an institutional product with a pace, a cycle, and a momentum of its own. Work on a new version again was begun before the preceding one was mailed to users. Using the now established procedures, the Working Group had drawn up a list of potential enhancements which was reviewed by the Management Group and given to Dr. Lindberg for final review. The highest priorities were: making more databases searchable via **GRATEFUL MED,** adding a search "edit" capability, and implementing an auto-install feature. In addition, Dr. Lindberg and the contractor worked together to design what came to be known as the **GRATEFUL MED Search Engine,** a software routine that allows system developers to incorporate the **GRATEFUL MED** search and retrieval capabilities into their own PC applications. The Search Engine has particular application for researchers from the hospital and academic community working under an NLM contract on the Library's ongoing Unified Medical Language System (UMLS) project.

During the development of this version, database experts from **DIRLINE, TOXLINE** and **TOXLIT, CANCERLIT, CHEMLINE, HEALTH PLANNING & ADMINISTRATION** and **AVLINE** were added to the Working Group to contribute and test the implementation of their particular files. Having so many extra hands led to a mistake: when it came time to Beta test, in-house testing was thought to be sufficient. Although sixty hours of formalized testing were done, there have been errors discovered in Version 3, which in retrospect would more likely have been found in outside Beta testing.

This was an especially hard-learned lesson because the Version 2 Beta test had revealed some equipment incompatibility problems. A lesson from the past had been ignored.

*Lesson #18: Don't assume that Beta testing can be done in-house. A wide variety of users and equipment and an extremely high volume of searches are imperative.*

The only delay in meeting a mid-December distribution date (to
coincide with the annual update of MeSH terms) again involved printing. Sixty percent of the User's Guide pages contained new material, so it was decided to reprint the entire manual rather than require the user to replace pages. The printing process caused only a three-week delay this time; Version 3 was distributed in early January 1988, ten months after Version 2.

By now it was clear that the Version releases are inexorably related to NLM's MEDLARS annual releases.

**Improvement by Analysis**

Up to now, improvements have been based primarily on user feedback, within the constraints of the design philosophy. That method of operation will be maintained; in addition, a more systematic assessment of the effect of a change will be made before implementing it.

To do this, a PC software log program has been written that collects data on how the user interacts with the software and the results of the search, but does not include data on the actual search content. Data on searches done in NLM's Reading Room with the current version will be collected and compared with data collected from a test version that contains a proposed change or enhancement. If the change does not produce the contemplated result, e.g., ease of use or improved retrieval, its implementation will be seriously questioned. Data collected for 5,000 searches in the Reading Room show that only 6 percent of the users looked at the Help screens; 20 percent used MeSH terms in their subject searches; and 40 percent of the searches retrieved nothing. This problem of zero retrieval will be the first to be approached with the new test stations. Nothing in the literature reports what fraction of the time searches generally retrieve nothing, but 40 percent is at least 25 percent too high. Significantly, this represents a pattern totally different from that seen for years with searches mediated by a medical librarian. These tend to have the opposite fault: that is, initial retrieval of far too many citations. There are various ideas for solving the new problem; each will be tried and data collected on the resultant change.

Both full function and test versions will be available side-by-side in the NLM Reading Room and at the National Institutes of Health Library, which is heavily utilized by clinical researchers. Consideration will be given to adding other test locations as this method evolves.

At this point in the development, the goal is to make the software "smarter" and more responsive to the end user's problems with searching. Plans are underway to construct a "hook" to an expert searcher program residing outside GRATEFUL MED. The results of the user's search attempts will be available to the expert program for analysis
leading to suggestions for improving the search, particularly in the use of vocabulary.

Other Evaluation Strategies

NLM has three evaluation projects underway that will undoubtedly influence the future development of GRATEFUL MED. A nationwide survey of nearly 3,000 health professionals who search the MEDLINE database, either by command language or using a front end such as GRATEFUL MED, is still in the data analysis phase; but preliminary results have underscored the importance of MEDLINE searching for patient care—69 percent indicated patient care as a primary purpose of their searching. This reinforces the goal of designing a system that best serves the information needs of the clinician.

A more elaborate study to collect information from health professionals about their use of MEDLINE using the Critical Incident Technique (CIT), an evaluative methodology that systematically collects and analyzes reports of users' actual behavior, is just beginning. In simplest terms, the CIT is used to determine critical requirements that have been demonstrated to make the difference between success and failure in carrying out an important part of a task. The goal is to understand and document how MEDLINE information is used, especially in patient care, and with what effect(s). That is, does the use of MEDLINE make a difference? The study results will be applied to improving the design of both the command language system and GRATEFUL MED.

A third project now being planned is to make use of a laboratory facility where users' behavior can be systematically observed under controlled conditions. Tasks will be designed that will provide information on how test subjects use the software and documentation while observed by "usability lab" staff. Data will be collected by recording a user's behavior/comments in a real-time computer log and by video-taping the interaction for later analysis. A GRATEFUL MED tutorial now under development is a likely candidate for this type of evaluation.

CONCLUSION

Rapid changes in micro and large computer technology and telecommunications capabilities, as well as the increasing level of computer sophistication in the health care professional community, promise ever greater levels of utilization of computer-based information resources for the nation's health care. Working on the design, development, and evaluation of GRATEFUL MED has been exciting and satisfying. But it is just a beginning.
ACKNOWLEDGMENTS

The authors wish to acknowledge the following persons who so generously contributed to the successful testing of GRATEFUL MED.

The following site coordinators participated in Phase II of the testing: Arthur Downing, New York Academy of Medicine; Frieda Weise, University of Maryland; Phyllis Self, University of Illinois at Chicago; Marie Reidelbach, University of Nebraska; Tricia McKeown, University of Texas at Dallas; Kay Denfield, University of Washington; Julie Kwan, UCLA; R. Brian Haynes, M.D., Ph.D., Ann McKibbon, and Lynda Baker, McMaster University; Joyce Mitchell, Ph.D., University of Missouri; Eleanor Goodchild, University of Pennsylvania; and Mary Horres, University of California, San Diego. Thanks are also due the following individuals who volunteered as end users: Randolph Miller, M.D.; Morris Collen, M.D.; Alfred Fishman, M.D.; Douglas Brutlag, Ph.D.; and Richard Friedman, M.D. In addition, Brian Haynes and Ann McKibbon of McMaster University and Julie Kwan from UCLA participated in the final Beta test group. New participants in this group were: Commander Joseph Henderson, M.D., Uniformed Services University of the Health Sciences; Phyllis Soben, Cedars-Sinai Medical Center, Los Angeles; Frances Brahmi, Indiana University; Barbara Rapp, Ph.D., Catholic University; and Gordon Black, Ohio State University. The Beta test group also included individual users Carolyn McHale, Donald Hawkins, and Andrew Dean, M.D.

REFERENCES