Application of Information Technology ■

Managing the Electronic NIH-Guide for Grants and Contracts

P. R. SMITH, MD, PhD, S. GOTTESMAN, MS, WILLIAM K. JONES, MBA

Abstract This article describes the implementation of a suite of computer programs to manage and provide access to a database containing the electronic documents that constitute the NIH-Guide that is distributed by the NIH on a weekly basis. The software consists of a management program that reads, processes, and stores the incoming documents and performs erratum updates on existing documents; an alerting program that sends selected information to users who have registered their information needs; a viewer that can be used on the local computer to read these documents; and a World-Wide-Web (WWW) server that can distribute the guide to computers that run WWW client software. The design of the documentation annotations, the management software, and the WWW server are expected to constitute valuable models for similar projects in the future.

■ *J Am Med Informatics Assoc.* 1995;2:94-101.

The National Institutes of Health (NIH) has the responsibility to guide and support the biomedical research community, which relies heavily on grants that the NIH offers. One of the tools that the NIH uses to communicate with this community is the NIH-Guide for Grants and Contracts, a weekly publication that contains postings of new regulations concerning the conduct of research and advertisements for new and continuing programs of research grant support.

Currently, the NIH-Guide is (primarily) a paper document, printed each week by the NIH in Bethesda, Maryland, and mailed to about 34,000 individuals and institutions throughout the United States and overseas at a cost of about one million dollars per year. Transmission of the paper copy via U.S. mail means that the guide is received at some institutions as much as three weeks after the publication date. Such delays create a significant problem for those recipients who need to examine it promptly. At the other end of the "interest" spectrum, a significant

number of these copies of the guide are thrown into wastebaskets unread.

Faced with the high cost of maintaining the production of the printed NIH-Guide and the frequent delays in delivery, the NIH began to investigate alternatives to the expensive, wasteful, and often ineffectual information distribution that it represented. Early testing showed there were primarily two types of individual readers of the NIH-Guide: those who waited for each issue and quickly examined it for grant possibilities, and those who looked at it occasionally or when they were looking for specific information. The printed guide met the needs of both of these types of individuals poorly, due to either the delays in delivery or the tiny fraction of information relevant to the individual reader. The only "consumers" of the printed guide that made good use of it were institutional grants offices, for which it was an essential publication.

The ingenious alternative to the paper publication was to use the growing availability of computer networks to provide a means of "publishing" the NIH-Guide as an electronic document. Accordingly, the NIH began a pilot project with a small number of institutions to test the feasibility of a purely electronic distribution of the guide. This was based on a LIST-SERVER, which provided a means to distribute the guide via e-mail while maintaining a backup copy that could be accessed if the e-mailed copy were lost or damaged. The original concept was to make one

Affiliations of the authors: Research Computing Resource, Department of Cell Biology, NYU Medical Center, New York, NY (PRS, SG); and Division of Computer Research and Technology, National Institutes of Health, Bethesda, MD (WKJ).

Correspondence and reprints: P. R. Smith, MD, PhD, Department of Cell Biology, NYU Medical Center, 550 First Avenue, New York, NY 10016. e-mail: <smithp01@mccr.med.nyu.edu>

Received for publication: 10/05/94; accepted for publication: 11/10/94.

electronic copy available to each institution. The institution would then use existing local mechanisms (LISTSERVs, bulletin boards, e-mail, etc.) to make the copy available to its researchers and administrators. A meeting was held in Bethesda, Maryland, in September 1990, to discuss the results of the preliminary test with administrative officials from numerous institutions, to collect suggestions for improving the distribution process, and to encourage institutions to subscribe to the E-Guide with a view to eventually phasing out the printed guide.

This article documents the response of the New York University (NYU) Medical Center to the NIH initiative, which was to write a database management and information display program package to process the weekly annotated feed from the LISTSERV, and to make it available to faculty and administration. The basic software package for OpenVMS/VAX was completed in November 1990, in time for the first general distribution of the E-Guide. It has been upgraded continually since its first release, and is employed at several other institutions as the principal data-processing package for these data.

We also document our response to the explosion in use of the World-Wide-Web (WWW),^{1,2} which has made a fundamental difference in the way the Guide information can be distributed. To improve the delivery of these data to our faculty, we have developed a set of server programs to allow the E-Guide documents to be served via WWW. This effectively solves the problem of cross-platform compatibility for people who wish to access these data since WWW client software, such as NCSA Mosaic,³ is available for nearly all common computers.

In this article we describe the annotation scheme used to prepare the data for e-mail transmission, the algorithm to process the incoming information and create the database, the management of the alerting software, the guide reading program used to provide local users with access to the data, and the development of the WWW server to permit access via the Internet. General issues concerning the distribution of databases are discussed.

Methods

The Structure of the NIH-Guide

Most of the information in the guide falls into one of three main categories: "Notices," which contain information about policy changes, documentation, meetings, and other administrative issues; "Requests for Proposals/Requests for Applications" (RFPs/RFAs), which are invitations to submit grant proposals or

contracts for competitive review; and "Program Announcements" (PAs), which describe ongoing programs for which institutions can apply. An additional category is "Errata," where updated information and corrections to previous Guide items are presented. The final component of the guide is the index, which lists every item and sometimes also contains additional information such as application deadlines. PAs and RFPs/RFAs can have full and short texts. Short texts provide the basic information and are included as a part of the printed guide. Full texts, which contain complete information about the RFP/RFA or PA, are distributed as part of the E-Guide but have to be requested separately by subscribers to the printed version.

It is important to be aware that the items that constitute the NIH-Guide are dynamic documents that can be changed each week. Although errata are not required very often, they are not uncommon and some documents are modified multiple times in successive weeks.

NIH-Guide Annotations and the Creation of the E-Guide

For the E-Guide to be processed by a computer program for insertion into the local database, an annotation scheme is required to demarcate components in each block of text sent via e-mail. These annotations are used to extract these components as separate files, to reassemble component fragments, to index these components so that the guide-reader and WWW server software can find them for display, and to apply the changes indicated by the errata. Although the goal of the E-Guide project is to create electronic documents that will eventually completely replace the conventional printed guide, it is necessary to be aware that many potential users do not have access to E-Guide-reading client software packages. Therefore, a major development criterion was that the annotated guide should be reasonably easy to read as a text document.

The annotation scheme as employed today is described in the appendix. In 1990, when the service was started, the annotation scheme was somewhat fragile, with many errors being noted in the processing the information received. Close collaboration between the NIH and NYU, with input from many users of the software, allowed the "bugs" to be eliminated from the annotation scheme and the guide-reading program that maintains the local database.

The weekly NIH-Guide is created by the NIH staff in a two-step process. First, the guide is created using

Table 1 ■

Procedures for Testing the World-Wide-Web (WWW) Server, the NIH-Guide-Reader for Open VMS, and the NIH-Guide Alerting Registration Form

1. Test the WWW server.
Open the URL: <http://www.med.nyu.edu/nih-guide.html>
2. Test the NIH-Guide-reader for OpenVMS.
 - a. Connect to "mcrcr.med.nyu.edu" via Telnet.
 - b. Login using the account named INFORMATION.
 - c. Select item 2 on the menu.
 - d. Read the HELP using the "H" key, exit the program using "E."
3. Test the NIH-Guide Alerting Registration form.
Open the URL: <http://www.med.nyu.edu/alert.html>

URL = Universal Resource Locator.

standard word processing software and printed for distribution to the U.S.-mail subscribers. As a second step, this word processor document is converted to a flat ASCII file; headings, footings, and white space are removed; and special tags are added that delimit the various components of the guide and create its electronic index. An "annotation" program then completes the annotation process and e-mails the fully annotated E-Guide from the NIH to the NIHGD-L BITNET list maintained at Johns Hopkins University. The main body of the guide and each full-text RFP/RFA are transmitted as separate mail items. Additionally, if any item exceeds 1,500 lines, it is broken into subitems of 1,500 lines or less.

The Guide-formatting Program

The Guide-formatting program processes the E-Guide when it arrives via e-mail each week in randomly ordered fragments. The first task is to extract these mail fragments and place them in a file for further processing. Accompanying the E-Guide information feed there may be "mail" for the local manager of the guide: this is processed separately and forwarded to that individual. The file containing the extracted guide data is processed to reassemble fragments and create individual files for each of the documents in the guide, including the full-text documents. Each document contains keyword, program type, and target group information. These terms are extracted and stored separately so that users can search the guide database using them.

Finally, the program can optionally archive guide items older than a specified time period, typically three months. The archival step takes the text files and places them in a compressed archive, from which

they are to be extracted automatically by the reading program. The archiving does not restrict the user's access to all parts of the guide, but it does make this access a little slower for the older items, since they must be decompressed before they can be read.

The structure of the database created by the guide-formatting program is very simple, with all data stored in several sequential files, which permits these files to be shared by other machines. We use this in our implementation of this software with the database's being shared via NFS between OpenVMS computers, which run the guide-reader software, and a UNIX workstation, which runs the WWW server.

The mail delivery and database update process using the guide-formatting program has been made very reliable and resistant to minor errors. The program runs entirely automatically on a weekly basis, with no intervention required. However, if the program detects errors, it generates a warning message that is e-mailed to the systems personnel for remedial action.

The Guide-reading Program for OpenVMS

The central component of the package is the user interface, which allows the user access to the guide items. The most important design criterion for this program was that it should be as intuitive as possible, allowing a user with no experience with the program, or even with computers in general, to navigate his or her way through it. Accordingly, the program was built using a menu-driven interface with single-key commands and extensive on-line help. It was implemented on VAX and AXP hardware from DEC (Digital Equipment Corp., Maynard, MA) under the OpenVMS operating system.

When the guide-reading program is used to browse the database, the user selects an issue of the guide from the main menu and then hits a key to choose the guide listing he or she wishes to inspect (RFPs/RFAs, PAs, or Notices). The selected list is presented as a menu, and the user can select to read either the short text or the full text associated with each item. Alternatively, if a cursory inspection of the whole guide is needed, the user can select the "full index" option and read through it. At each step in the menu-selection process, the user has the option of extracting the information presented on the screen to a file, e-mailing it to another user, or printing it locally or to a system printer. This means that specific guide items, guide indexes, or an entire guide can be obtained in any form that is appropriate.

The functionality of the software can be tested by following the procedure outlined in Table 1.

The WWW Server Program

The most important change in the distribution of the guide recently has been the observation that the E-Guide database can be served to the Internet using the WWW protocols. In our implementation (using a DEC 3000 Model 700 AXP workstation running OSF/1), the WWW server presents the user of a WWW client with a document, updated weekly, with links to all the guides in the database, to guide-searching tools, and to the Guide Alerting Registration form. From a functional standpoint, the WWW server offers all the services provided by the local guide-reading program.

If the user selects a particular issue of the guide from the list, the next screen he or she will see is the full index to that guide, and individual items can be selected from this list. If that guide contains erratum items, these will be presented individually and will contain links back to the items they modify (which have already been updated with the new information). When the server receives a request with the selection from the user's WWW client, it responds by building an "html" (HyperText Markup Language²) file on the fly from pieces of textual information taken from the database. A server program is required because the WWW server and the local guide-reader program share the same database and also because the server has to be able to include time-sensitive information.

The guide-searching form allows the user to find guides that have been indexed with the keywords from the list that the NIH uses for annotation. The Guide Alerting Registration form allows a user to add his or her e-mail address, password, and a list of keywords to the alerting database.

Our experience with the use of the WWW server software has been very positive, and those interested can test it directly following the procedure in Table 1. Logs from the WWW server show, on average, access to approximately 3,000 guide documents every month, with a steady increase being noted.

The Guide Alerting Program

One of the major benefits of computer processing of the guide is that it offers the ability to automate some aspects of the information delivery process. One important issue is to ensure that faculty are properly made aware of the support available to them in the research areas where they work.

We have taken a first step in this process by creating an "NIH-Guide Alerting" component of the guide-formatting program. This program creates a list of

the keywords used to index incoming guide items. It then compares the list with a list of keywords linked to e-mail addresses of faculty who have requested the service using the WWW Guide Alerting Registration form. If any of the keywords for the incoming guide items matches keywords in an individual's list, an e-mail message is sent to that person, alerting him or her to the arrival of that guide and providing full information, including a Universal Resource Locator (URL),¹ which points to the text of the item. This scheme obviates the necessity of mailing the entire text of the guide item since the URL allows direct access to the relevant information.

Establishing a Local NIH-Guide Database and a WWW Server

All software developed as part of this project is available at no charge to those who would like to have it. The process for downloading this software package, including all the information needed in order to establish this service from scratch, is described in Table 2.

In addition to the software, an appropriate hardware platform is needed with disk space to store a local database containing the NIH-Guide information. At NYU, we use computers running the OpenVMS operating system; but in the V1.3 release of the software, support for UNIX is provided. The software distribution describes in detail how a site can establish its own WWW server and alerting databases. Since distributions contain full sources, each site can customize the software to meet local needs.

Table 2 ■

How to Download and Obtain the Software Used in This Study

1. Those interested can download the software to their own machines using the MAILSERVer at mcrcr.med.nyu.edu.

Prepare an e-mail message to the address MAILSERV@mcrcr.med.nyu.edu. In the body of the message place the words HELP and INDEX on separate lines, then send the message. The reply you receive will contain instructions for using the MAILSERVer and a list of the packages currently available. The NIH-Guide package is currently available in files with names like NIH-GUIDE.xx-OF-10; however, these names change as the package is updated. Please use the INDEX, which provides up-to-date information. More extensive help for using the MAILSERVer is available from the URL: <http://www.med.nyu.edu/mailserv.html>.

2. The software can be ordered from the DECUS Software Library, part number V00557. DECUS can be reached at 1-800-332-8755; please specify V1.3 as the minimum release.

Although the use of the WWW is growing rapidly, faculty and administrators need to be made cognizant of its power. At NYU, the medical library provides training and consultant services for the faculty concerning information technology, and the grants office and the library cooperate in the dissemination of information about the NIH-Guide and electronic access to it. Our experience is that minimal support is needed for the NIH-Guide software; nevertheless, sites wishing to establish their own databases and WWW servers need to plan to inform the potential users so that the maximum benefit can be obtained from this new local resource.

Discussion

Managing Electronic Documents

This article presents a solution to a specific problem related to two issues of growing importance: 1) the need to acquire and store information as a resource for individual and institutional tasks, and 2) the need to identify key pieces of information that have an immediate impact on the way tasks are performed and plans for the future are formulated.

Our software permits a solution to the problem of increasing the efficiency and effectiveness of the NIH-Guide. With the use of computer networks, we have ensured that the guide can be received electronically in most institutions. With the development of the annotation scheme, we have allowed the information in the guide to be processed automatically and placed in a local database with an efficient mechanism for access by guide, item, or keyword. With the institution of an automatic erratum-update procedure, we have ensured that the information extracted, mailed, or printed from the database is the most accurate available. The alerting feature ensures that individuals are informed immediately when any item relevant to them is modified or is added to the database.

This project is closely related to earlier work by our group,⁴ in which we designed and built software to maintain an up-to-date copy of the GenBank⁵ (the database that stores the totality of DNA sequences discovered thus far) from nightly updates made available via USENET. A similar package has been developed to provide local updating for the EMBL DNA database.⁶

The use of computers to provide automatic alerting of individuals to situations of importance to them in the biomedical setting is growing rapidly. For example, clinical decision support systems are under

development to assist in disease diagnosis, management of patient data, and alerting clinicians to abnormal conditions.⁷ In an application being developed at Yale University, the Current Contents database is linked to an alerting service based on electronic mail.⁸ Our most recent release of the GenBank updating software (available from our MAILSERV) also contains an alerting feature based on a package described by Jones⁹ that allows subscribers to this service to specify which sequences or classes of sequences should be e-mailed to them if found in the incoming sequences data stream. These examples of alerting software facilitate the clinician's and the researcher's task of screening large volumes of information for critical items that can be identified using rules programmed to permit computer identification. They therefore help protect the integrity of the clinical or research process.

Growth in the Use of the E-Guide

The NIH E-Guide was one of the first documents specifically designed for electronic distribution. Its acceptance was slow. In 1990, scientists and administrators, even at research institutions, were not comfortable reading documents electronically, at least in part because there was virtually no software available to assist with the task.

Examining documents electronically is now becoming a way of life for individuals who have access to the Internet. Indeed, the benefits of networking for information access are being promoted in publications such as *PC Magazine*, where a good guide to network tools has been presented,¹⁰ and *Science*, which recently focused on networking issues, including information access.¹¹ This change in the culture of the general user is reflected in the fact that the LISTSERV now has over 2,500 subscribers. It has allowed the NIH to take the next step and announce that the primary distribution mechanism for the NIH-Guide will be electronic. Accordingly, the NIH has restricted the distribution of the printed guide to institutions, and individual subscriptions were eliminated effective August 1, 1994.¹² Individual subscribers are now forced to employ electronic means to access these data.

Provision of the Database Service

Given that the creation of a database containing these data solves the problem of effective delivery of the guide, the question then is, what kind of database should be used? Should it be on a centrally located server (in Bethesda, Maryland, for example), or should the servers be locally maintained?

At the time the E-Guide project was started in 1990, most interconnections between university sites were via slow network links. Even discounting the high cost to the NIH of providing computational power to support interactive access to a central database for an NIH-Guide browser, the network links to all but a few institutions would not have supported it. The only viable option was to create a properly annotated version of the guide for e-mail each week via a LISTSERV, where backup versions of old guides could be preserved. The E-Guide was born and institutions were encouraged to subscribe to it and to find ways to make access available locally.

A locally maintained database updated weekly via e-mail offers many benefits. First, its integrity can be checked and errors can be repaired locally, thus providing a high level of quality ensured by institutional efforts. Second, the availability of the database is under local control, allowing access to be independent of computer and network outages external to the institution. Third, access to the database can be monitored, allowing specific uses of the information to be tracked. Finally, the institution can offer alerting software to forward information to the individuals best able to respond to it. In the past, a disadvantage of a local database was the cost of maintaining it, including both the time of the manager and the disk space to hold it. In operation at NYU, however, the database has proven to be almost a "no-cost" item. The aggregate size of the database, since its creation in 1990, is about 40 Mb. The cost of a high-speed disk to hold these data is less than \$30, based on the present prices for disks, which average \$650 per gigabyte. Management, from the system manager's perspective, has been miniscule since updating the database is entirely automatic.

Even so, in 1995, the benefits of a centralized, nationally maintained database are more attractive. Interinstitutional trunk network links are now much faster and significantly more stable. The distribution of the guide using the WWW, a low overhead protocol, makes large numbers of remote database queries more manageable. It is feasible to serve a large region without dedicating a machine to this task alone. Central management eliminates concerns over the reliability of the e-mail data feed and also makes the guide available about a day sooner than it is at remote sites, although neither of these issues has been problematic for the user community thus far.

Other Ways to Access the E-Guide

Although our software has found a home in several institutions, its use has been limited to sites that have access to computers running DEC's VMS operating

system. Some sites have written their own internal distribution systems based on the LISTSERV feed from Johns Hopkins University, including providing a local LISTSERV for their faculty. Other sites have their faculty make direct use of the raw, annotated guide via USENET news since the LISTSERV is gatewayed directly to the "bionet.sci-resources" newsgroup, which also carries the NSF's Science and Technology Information System (STIS) summary of the documents added to the STIS system in the previous week.

The NIH has responded to the need to facilitate electronic access to the E-Guide by opening the LISTSERV to general subscription and by providing guide information via the NIH-Gopher and the NIH Grantline bulletin board. The NIH is also creating a new service called GrantsNet,¹³ the mission of which is "to serve as a vehicle and catalyst for continuous improvement and innovation in federal grants management practices, policies, and information dissemination."

The problem with the mechanisms presently in place (or presently described as future services of GrantsNet) is that they do not permit the use of all the information provided in the E-Guide annotations. While it is quite possible to write a Gopher server to manage these dynamic aspects of the E-Guide in a fashion similar to the WWW server we described here, this has not been done as yet. Users of the E-Guide as raw data from the LISTSERV, from USENET, or from the NIH Gopher (or the NIH WWW server, which currently points to the NIH Gopher) must therefore ensure that the data they have retrieved are complete and up-to-date by inspecting all subsequent guides for posted errata.

Our development of a WWW server for the E-Guide database solves these problems. Our server software fully implements all information provided in the annotations and document headers; therefore, it provides a complete and up-to-date view of this document at the time the information is accessed. We are unaware of similar software available today that provides the full range of these services.

Conclusion

The "information explosion" is creating significant problems that touch everyone in society. In the home it is manifested by waves of unsolicited advertising material, in the office it is reams of administrative memos, reports, and catalogs, and for the scientist it is the burgeoning number of scientific articles. The information that the professional researcher needs to

have available is increasing in volume, both because there is more being generated and because of the speed at which this information goes stale as new regulations are promulgated, new equipment becomes available, and new techniques are discovered that become the "standards of business."

The NIH-Guide is a small piece of the huge information infrastructure required to support academic research. With the advent of the WWW, and particularly the graphic WWW clients such as Mosaic³ that can use it, the expectation is that services like this will grow exponentially with fundamental consequences for the delivery of many types of information to the academic community.

In the last five years, library card catalogs increasingly have been replaced with text-based computer terminals and have been turned into electronic catalogs. These text-based services have often been extended by adding links to Gopher client software, effectively extending the catalog with links into libraries worldwide. The emergence of internet-worked, hypertext-based information resources with their intrinsic multimedia content heralds a new paradigm for the library. From a "library without walls" we are moving toward "information without libraries," at least in the form that most people currently recognize them. Indeed, discrete repositories of obsolete information duplicated hundreds or thousands of times in libraries worldwide has the potential to be a barrier to information access, rather than to facilitate it. Some of these duplicated resources will be eliminated entirely with time. For the remainder, software and protocols that permit local databases to be kept synchronized with the central information source, similar to the erratum-update service we have implemented, will become increasingly important.

Our development of the NIH-Guide-reader software has occurred at the cusp of considerable change in the way in which information is created and distributed. When we began in 1990, the existence of discrete repositories of these data was a necessary and inevitable response to the technology available. Today, in 1995, we can foresee a time when a local

NIH-Guide database may become obsolete. We hope to remain effective contributors to the evolution of information technology through new developments to this and other related software.

The authors thank John Hill, Foteos Macrides, Frieda Pavel, Loren Buhle, Claudia Blair, and John-Paul Elrod for their aid at several stages of this project. NYU's NIH-Guide database management software was developed on a VAX 6000-410 computer supported in part by a grant from the National Science Foundation (DIR-8908095 to Dr. Smith). The WWW server software was developed as a project of the NYU General Clinical Research Center, which is supported by an NIH-NCRR grant (RR-00096-32 to Saul Farber).

References ■

1. Berners-Lee TJ, Cailliau R, Groff J-F, Pollermann B. World-Wide-Web: The Information Universe. Electronic Networking: Research, Applications and Policy. Westport, CT: Meckler Publishing, 1992:52-8.
2. Berners-Lee T, Cailliau R, Luotonen A, Nielsen HJ, Secret A. The World-Wide-Web. *Comm ACM*. 1994;37:76-82.
3. Schatz BR, Hardin JB, NCSA Mosaic and the World Wide Web: global hypermedia protocols and the Internet. *Science*. 1994;295(5174):895-901.
4. Smith RH, Gottesman S, Hobbs B, et al. A mechanism for maintaining an up-to-date GenBank database via USENET. *Comput Appl Biosci*. 1991;7(1):111-2.
5. Benson D, Lipman DJ, Ostell J. GenBank. *Nucleic Acids Res*. 1993;21(13):2963-5.
6. Williams GW, Gibbs GP. Automatic updating of the EMBL database via EMBNet. *Comput Appl Biosci*. 1990;6(2):122-3.
7. Frost RD, Gillenson ML. Integrated clinical decision support using an object-oriented database management system. *Methods Inf Med*. 1993;32(2):154-60.
8. Paton JA, Belanger A, Cheung KH, et al. Online bibliographic information: integration into an emerging IAIMS environment. *Proc Annu Symp Comput Appl Med Care*. 1992:605-9.
9. Jones R. Alerting users to relevant new entries in the GenBank DNA sequence database. *Comput Appl Biosci*. 1992;8(2):199.
10. Ayre R. Making the Internet connection. *PC Magazine*. 1994;13(17):118-39.
11. Waldrop MM. Software agents prepare to sift the riches of cyberspace. *Science*. 1994;265(5174):882-3.
12. The NIH Guide for Grants and Contracts: Intent to Modify. The NIH-Guide. 1994;23(23):Notice #1.
13. The NIH Guide and Contracts: Development of GrantsNet. The NIH-Guide. 1994;23(38):Notice #3.

information that the professional researcher needs to be able to access is being generated and distributed in a way that makes it possible to review by computer in the office. It is time to consider the need for a central repository of scientific information and for the development of a system that can manage this information in a way that makes it possible to review by computer in the office. It is time to consider the need for a central repository of scientific information and for the development of a system that can manage this information in a way that makes it possible to review by computer in the office.

Although our software has found a home in several institutions, its use has been limited to those that have access to computers running DEC's VMS operating system.

APPENDIX

Annotation Scheme as Employed Today

The annotation strategy is as follows. Each main document, the E-Guide itself, and each full-text document is introduced by an ID line that has the following format:

```
$$$ID file-name file-type file-descriptor sequence-id
```

For the E-Guide, the file-name is NIHGUIDE, the file-type is the date of the guide in the format yyyyymmdd (e.g., 19910809), the file-descriptor is the volume number and the issue number in the format VnnNnn (e.g., V12N06), and the sequence-id, which is used to identify parts of a single document whose lengths exceeded 1,500 lines and therefore had to be sent separately, is in the format PnOn (e.g., P2O3—part number 2 of the 3 parts that were sent). An example of the ID line for the main body of an E-Guide is:

```
$$$ID NIHGUIDE 19910809 V12N06 P203 *****
```

For a full-text RFP/RFA, the file-name is RFP/RFA, the file-type is the RFP/RFA number, excluding any nonalphanumeric characters (and therefore not exceeding 8 characters in length), the file-descriptor is the full RFP/RFA number. The sequence-number is the same as described above. An example of the ID line for the full text of an RFP/RFA is:

```
$$$ID RFP/RFA DA9112 DA-91-12 P101 *****
```

The main body of the E-Guide contains the index followed by the short text for each index entry. Before the first line, and following the last line of the index, these delimiter lines are found (respectively):

```
$$INDEX BEGIN and
```

```
$$INDEX END
```

Before each index entry, a line with the following format is found:

```
$$INDEX tn mm/dd/yy
```

where t is a letter indicating the text-type (R for RFAs/RFPs, N for notices, P for program announcements, and E for erratum) and n is the sequence number of that type of entry for that guide. If there is an expiration date, it will be given in the format mm/dd/yy.

Surrounding each short text in the main body of the E-Guide are the following lines:

```
$$tn BEGIN RFP/RFA/PA-number FULL-TEXT
```

```
$$tn END
```

The RFP/RFA/PA number is present for RFPs/RFAs and PAs but not for other short texts such as notices. The word FULL-TEXT is present if a full text of the RFP/RFA or PA has been distributed by the NIH.

Erratum entries in the E-Guide have the following format:

```
$$En BEGIN tn yyyyymmdd APPEND/REPLACE RFP/RFA/PA-number FULL/SHORT/BOTH
```

The n following the \$\$E is the erratum number in the present guide. The tn is the text-type and number of the original entry. The yyyyymmdd is the file-type of the original entry. Either APPEND or REPLACE is specified depending on whether the original entry is being modified or is being replaced by the text of the erratum. RFP/RFA/PA-number is the number of the original entry and the word FULL, SHORT, or BOTH is specified depending on whether the full text, the short text, or both texts are being modified.