Developing a subject-specific Gopher at the National Library of Medicine

By Gale A. Dutcher, M.L.S.
Special Assistant to the Associate Director

Stacey J. Arnesen, M.S.
Technical Information Specialist

Specialized Information Services
National Library of Medicine
8600 Rockville Pike
Bethesda, Maryland 20894

Development of subject-specific Gophers is one method to make valuable health information more accessible to those who need it. The National Library of Medicine has developed a Gopher that provides access to information produced by many areas of the library. Recommendations from subject experts in two areas—AIDS, and toxicology and environmental health—led to the development of two subject-specific Gophers. These two Gopher services provide organized access to resources outside of the library as well as to information produced internally. A number of important issues need to be addressed when selecting and organizing outside resources for inclusion in such a service.

INTRODUCTION

The National Library of Medicine (NLM) is involved in a wide-ranging project to develop a Gopher service as one method for making valuable health information easily accessible to those who need it. This Gopher service is only one component of NLM’s Internet services.

The precursor to NLM’s Gopher was a publications file-transfer protocol (FTP) server known as NLMPUBS. This was found to be a useful mechanism to easily distribute fact sheets, bibliographies, and other forms of printed information that otherwise would require mailing. Although providing FTP access does not eliminate mailing, it certainly makes it easy and fast for a large number of people to access the information. The Gopher evolved as the next level of service NLM could provide for users to locate needed information. The Gopher essentially uses the same directory structure as NLMPUBS and provides access to the same files.

The NLM Gopher provides access to information produced by many areas of the library. However, additional development has gone on in two specific subject areas—AIDS, and toxicology and environmental health. The Toxicology and Environmental Health Information Program (TEHIP) and AIDS Gophers not only identify NLM resources, but also point to relevant resources available from other sources, such as descriptions of courses and meetings, clearinghouses, and other Gophers. These two Gophers are not physically distinct from the NLM Gopher, but they are logically distinct.

RATIONALE FOR DEVELOPMENT

The impetus for the development of both the AIDS and the TEHIP Gophers came from recommendations by experts in these fields. Specifically, the TEHIP Gopher was implemented in response to the report of the Long Range Planning Panel on Toxicology and Environmental Health, Improving Toxicology and Environmental Health Information Services, which was issued in September 1993 [1]. Goal 2 of that report states that NLM should “Facilitate access to national and international information resources for toxicology and environmental health.” The report also includes a specific recommendation about organizing resources, recommendation 2.2: “Develop a directory for toxicology and environmental health information resources.”

The development of the AIDS Gopher was initiated in response to discussions that occurred at the National Institutes of Health (NIH) HIV/AIDS Information Services Conference, held in June 1993 [2]. Participants at that conference made a number of rec-
ommendations concerning the complexity of accessing HIV/AIDS information and the need to simplify this process (e.g., recommendation 3.5: “NIH should work with other involved agencies to organize information resources to make it easier for allied health care providers to access information”; recommendation 5.12: “NIH should expand its AIDS information services available on the Internet . . .”). In addition, a forum entitled “Future Directions in AIDS Research,” cosponsored by the Harvard AIDS Institute and Project Inform (a not-for-profit AIDS advocacy organization) included a working group that looked at electronic communication. In its unpublished draft report, this group specifically recommended, “The National Library of Medicine should collaborate in the development of an electronic Gopher system that is specific to the needs of AIDS research, for both basic scientists and clinical researchers.”

Gopher technology was chosen because of its ready availability to the library and its widespread use in the community. The staff also was aware of a large number of useful Internet-accessible resources available in these two subject areas. This type of Gopher development was a departure from the general philosophy of Gopher development at NLM, because the general NLM Gopher includes only NLM-produced resources. The TEHIP and AIDS Gophers go further and serve as directories to additional Internet-accessible resources in these fields.

METHODOLOGY

NLM resources

Selecting NLM resources for inclusion in the Gophers was relatively straightforward. For example, fact sheets about the various TEHIP and AIDS programs and services are included in these Gophers. These fact sheets are also accessible from the main menu of the NLM Gopher. There is only one actual copy of each file, the various menus simply point to that file. This approach minimizes any problems that might be associated with updating the contents of these files while facilitating access to them no matter where entry into the Gopher is made.

In addition to the fact sheets, the Gophers include other support materials, such as users’ manuals for searching these databases on the NLM system, applications for user codes, sample search scenarios from several of the more complex toxicology databases, and NLM’s computer-aided instructional packages: TEHIP Databases Demo Program, ELHILL LEARN, MEDITUTOR, and TOXLearn.

NLM also has developed a number of resource lists, bibliographies, and reports that are available on the Gophers. These include the Guide to NIH HIV/AIDS Information Resources, AIDS Bibliography, Health Hotlines, and a calendar of meetings and courses in toxicology and environmental health. As new resources are developed, they will be added to the appropriate Gopher.

Non-NLM resources

Because the effort required to identify and organize external resources is a significant undertaking, two groups worked independently on this project. The NLM staff worked on identifying resources for the TEHIP Gopher, and NLM library associates identified those for the AIDS Gopher. Both groups took essentially the same initial approach. A very wide net was cast using the broadest definitions of scope to find a larger number of Internet-accessible resources. The Gopher name and address (host, selector string, port) were identified and collected. After a substantial number of resources were identified, there was a review process. A subset of the initial list was selected for final inclusion in the Gopher and organized into the Gopher menus.

SELECTION OF RESOURCES

What is the best way to decide which of the many resources found on the Internet should be included on the NLM TEHIP and AIDS Gophers? The main purpose of creating these Gophers was to simplify access to resources produced by others by providing a directory to these resources. However, there is likely to be some expectation by users of NLM services that these items have been reviewed and provide current and accurate information.

Scope

NLM’s Scope and Coverage Policy Manual provided some guidance in selecting resources for inclusion, particularly in the TEHIP Gopher. Although pointing to resources on the Internet is not precisely the same as acquiring them for the collection, the manual can provide some direction. In addition to selecting resources according to the scope described by the manual for adding to the physical collection, information about regulations and the basic chemistry of hazardous substances was sought. Also included, of course, was information on the essential topics of toxicology, pharmacology, occupational health, environmental health and medicine, risk assessment and management, health effects of pollution, and public health.

The Scope and Coverage Policy Manual provided less assistance in the area of AIDS resources. Although the AIDS Gopher focuses largely on medical and health aspects of AIDS and HIV infection, some of the Internet resources listed include topics peripheral to NLM’s scope and coverage policy. For example, there are numerous Internet resources that include information on housing, social security, health in-
urance, and legal information as they relate to AIDS and HIV infection. In addition, many resources refer to local projects. NLM's goals are to include primarily information of interest to a national and international audience and to focus on health-related issues. Therefore, a few sample or model local and non-health-related resources are described, but they are not covered comprehensively.

Review of information producers

As with selecting any type of material for a collection—and these Gophers may be considered collections of resources—who produces and authenticates the information is an important consideration. Information produced or approved by an agency of the U.S. government is regarded. Other governmental sources, such as state and non-U.S. government agencies, also are excellent sources of information. This is not necessarily because the information is better than nongovernment information but rather because of the standing of government agencies. It is assumed that the government agencies review and approve their data prior to distribution and that the information they make available publicly is accurate and represents the government's position.

There are also other authoritative sources considered for inclusion in these Gophers, such as the United Nations (UN); the World Health Organization (WHO); and professional societies, including the American Chemical Society. Important information is also available from other types of organizations, such as universities and voluntary agencies. Each of these groups provides its own type of review for the materials it produces or to which it provides access.

However, the actual producers and reviewers of many Internet resources are often not clearly indicated. Even if the information appears to be valid and covers an important topic, if no source is provided, it is difficult to justify including it in an NLM Gopher. For example, a nongovernment entity might download government information and make it available on the Internet. However, it may not be kept up to date, it may have been modified by that organization, or it may be taken out of context.

One mechanism that can be used to provide access to resources that appear to be significant but do not indicate the producer is to include a statement about that lack of verification so that users understand the lack of documentation. NLM staff may attempt to contact the source of the information if an electronic mail address is available in order to obtain more information about the resource. A number of the AIDS resources are provided by nongovernment and community-based organizations. NLM is not in a position to judge the validity of the information provided by these groups. However, the user community wants to have simple access to this information. Because these resources may not undergo the same strict review process that government-produced information does, this fact is conveyed to users in the documentation about the Gopher. It is most important that users not be misled about the validity of the data they may find by going through NLM's Gopher.

Stability and reliability

The two previously mentioned issues—scope and producer—should not be surprising to librarians because they apply to the selection process for book, journal, and audiovisual collections. Somewhat unique to the Internet environment is the issue of stability and reliability of access to the resource. There is a great deal of activity on the Internet, and resources are modified, moved, and removed with great frequency. A Gopher name or address may change, the menus may be rearranged, and files may be added or deleted. All of these kinds of changes lead to problems in providing access to these resources. It sometimes seems as though this is ephemeral information. Even more frustrating, there is often no easy way to identify a human contact to follow up with some of these problems.

Currency of information

Internet users expect up-to-the-minute information. Users expect the information superhighway to access the best, most recent information with a few keystrokes. Unfortunately, that is not always the case. Just like more traditional formats, Internet-accessible resources require effort to maintain their currency. Because widespread use of the Internet is still relatively new, the structures may not always be in place to keep resources up to date, and a number of groups providing access to their data may not understand the extent of the effort required to maintain a useful product. Many such files and other resources are not tagged with date information, so it may be difficult to ascertain how current the information is.

For the TEHIP and AIDS Gophers, NLM resources are always tagged with the date. If available, the date is included for the non-NLM resources as well.

ORGANIZING THE INFORMATION

Once useful resources have been identified, the question of organization must be addressed. For example, how fine should the divisions among Gopher topics be? Should the information be divided by type of resource, by producers, or by subject?

Menus

The menus presented to users are the public view of the organizational structure of the resource. Users may not be aware of the philosophy underlying the
menus displayed, but they know whether it is easy or difficult to find the needed information. Unlike classroom instructors, Gopher developers do not get an opportunity to explain their organizational philosophy directly to the users, although a useful way to convey this information is in a file on the Gopher often listed as "about this Gopher."

**Text of entries**

In developing Gopher menus, there are some constraints that differ from those that apply to catalog cards or bibliographic citations in an online database. To ensure that the entire line of text is visible to most users no matter what type of Gopher client is in use, each line of the menu cannot have more than sixty to seventy characters. Therefore, the menu entry has to be pithy. Because no additional information is readily available on the screen to provide clues to the user, it is important for the text in the menus to be as precise as possible to ensure that it clearly defines what is included in that menu choice. Otherwise, users will become frustrated and stop using the Gopher. There is clear conflict between the need to be as descriptive as possible yet brief in the Gopher menu entries.

One mechanism to help alleviate confusion is to include an "about this section" file for each menu. All menus in the AIDS and TEHIP Gophers include "about" files that briefly explain the choices in each menu.

**Organizational system**

A number of methods to group the information logically were reviewed and evaluated. These methods included separating NLM-produced resources from the external resources; separating resources by producer (U.S. government, other government, and non-government); separating resources by subject terms, such as Medical Subject Headings; and separating resources by format (files, Gophers, listservs, etc.). Combinations of these schemes also were examined. Organization by producer may be useful, because certain types of attributes may be ascribed to each producer category. Major topical areas such as chemistry, occupational health, and risk assessment also facilitate access for users.

In both the AIDS and the TEHIP Gophers, NLM information resources are separated from the non-NLM resources. It has proven simple to organize the internal resources in a logical way. Although each item included in the Gopher exists in only one physical entity, such as a file, it may be listed several times within the Gopher menu structure. By listing the item multiple times, users have improved opportunities to locate the item no matter how they approach their inquiry. The single physical entity mitigates the problems associated with maintaining the currency of the information.

Figure 1 shows the main menu of the TEHIP Gopher. The first item is a text file called "About the TEHIP Gopher Resource," which explains the purpose and organization of the Gopher. Items 2 through 5 describe various NLM resources. The NLM resources are listed first in both the AIDS and TEHIP Gophers. Item 6 is a description of events (meetings, courses, etc.) in toxicology and environmental health compiled by NLM staff. Item 7 will contain case studies on various chemical hazards. Item 8 contains a bibliography on alternatives to animal testing produced by NLM.

Non-NLM Internet-accessible resources are listed in items 9 and 10. Item 10 describes general Gophers, listservs, and telnet services dealing with toxicology and environmental health. In most cases, the TEHIP Gopher points users to the top or main level of a relevant Gopher, such as those run by the Environmental Protection Agency or the National Institute of Environmental Health Sciences. Item 9 lists more specific Internet-accessible resources than are included in item 10 and lists them by subject. Examples include WHO's Programme for the Promotion of Chemical Safety, the National Toxicology Program, and issues of *Morbidity and Mortality Weekly Report*. NLM has used a combination of approaches including producer, format, and subject to organize the subject-specific Gophers.

**Levels of menu entries**

Another issue is determination of numbers of Gopher levels, or depth of menus. How many levels are needed to adequately display the specific information the user is seeking without the user getting lost in the Gopher? Are a large number of items listed on one menu better than a few items on several layers of menus? To deliver the information with as much specificity as necessary, there may need to be a large number of items listed on each level of the menu.

In developing the menu structure for the TEHIP
and AIDS Gophers, the number of levels were kept
to a minimum. At this stage, none of the TEHIP Go-
pher menus are more than three levels deep. Figure
1 is the main menu of the TEHIP Gopher. Figure 2
shows the second-level menu that appears after item
5, “Documentation,” has been selected. Figure 3 shows
the final level of this sequence after item 3, “TEHIP
Search Scenarios,” has been selected from level 2.
Each item listed on this third-level Gopher menu is a
file presumably containing the information the user
is seeking.

Maintenance issues

To facilitate access to specific types of information, it
may be desirable to point to only a single file or entry
in another Gopher. In addition, the amount of infor-
mation available within a Gopher may be large enough
that the menus may need to contain five, six, or more
levels. For example, finding the UN document Science,
Risk Analysis and Environmental Policy Decisions on
the UN Gopher requires the user to traverse five levels
to find out where the desired document actually is
located (level 1: United Nations; level 2: UN Envi-
ronment Programme (UNEP); level 3: publications;
level 4: UNEP environment and trade monographs;
level 5: list of titles). This may be a very useful and
important document, but finding it through the Go-
pher may not be intuitive for all users.

In developing a menu structure for pointing to ex-
ternal resources, it is necessary to balance the ideal
of pointing to the specific item(s) identified as rele-
vant with the potential for their changing. The root
Gopher is a much more stable entity to include in the
Gopher menu but much less specific than the file of
particular interest. However, the overall external Go-
pher may not be specific to the topics highlighted.
Should the pointer be to the root Gopher, which is
not likely to experience frequent address changes, or
to the relevant subject-specific items within the Go-
pher?

To continue this as an ongoing operational service,
the list of external Internet resources will have to be
maintained. The maintenance plan for these Gophers
includes periodic verification of the external re-
sources included. In addition, “All the Gophers in

CONCLUSION

A great deal of important information is available on
the Internet. Gopher is a useful mechanism for or-
ganizing this information. It is a tool that is readily
available to a large number of users, including those
with dial-in access to the Internet as well as those
directly connected. A large number of Gopher clients
are available at no cost. Gopher software is relatively
simple to use for both Gopher creators and users.

Transforming a Gopher into an organized directory
of valuable information requires a significant amount
of time and effort, however. The two subject-specific
Gophers developed by the Specialized Information
Services Division of NLM promise to alleviate some
of the burden of locating valuable resources on the
Internet by providing current and evaluated direc-
tories of resources in the areas of toxicology and en-
vironmental health and AIDS.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the following
people for their work on the Gophers: Joyce Backus,
Cynthia Burke, Eileen Madden D’Andrea, Patti Dick-
inson, Jeanne Goshorn, Jennifer Grady, Carol Hab-
berman, Mike Hazard, Vera Hudson, Egle Locatis,
Dorothy Moore, Mark Silverman, Alice Su, Phil Wex-
ler, and Cheryl White.

REFERENCES

1. NATIONAL LIBRARY OF MEDICINE, BOARD OF REGENTS. Impro-
ving toxicology and environmental health information

Received October 1994; accepted November 1994