(2) The acronym "NSC" stands for National Service Center, a short form of Cancer Chemotherapy National Service Center, the early name for the program. The "NSC Number" is used by NCI as a Registry Number. Other Registry Numbers, such as the CAS Registry Number, are not useful for NCI because CAS Registry Numbers cannot be assigned to the confidential structures which constitute about half of the NCI database.
(3) These standards are outlined in "Guide for the Care and Use of Laboratory Animals", Institute of Laboratory Animal Resources, National Research Council. *NIH Publ.* **1985**, *No.* 85-23.
(4) The appropriate level of activity varies from one tumor system to another and depends upon whether animal survival or tumor inhibition is being measured. Activity criteria for all standard tumor systems are published by NCI, see: Geran, R. I.; Greenberg, N. H.; Macdonald, N. M.; Schumacher, A. M.; Abbott, B. J. "Protocols For Screening Chemical Agents And Natural Products Against Animal Tumors And Other Biological Systems", 1st ed.; *Cancer Chemother. Rep.* **1972**, *3*. The third and current edition of this report is published as NIH Publication 84-2635, available from the U.S. Government Printing Office, Washington, DC, 20402.
(5) Mice are not long-lived animals, and consequently all the screening laboratories must be supplied with mice at a rate and in quantities which match the flow of chemicals. Assuring an appropriate match is a major task, and it is not undertaken by the DIS. Rather, an independent computer system, loosely articulated with the DIS, manages the flow of animals.
(6) A total of 52 standard vehicles is permitted in animal screening, and this group is, in fact, one of the minor DIS databases.
(7) Full details of all screening procedures are published by NCI in "Instruction 14", 1985. Copies of this booklet may be obtained from the Information Technology Branch, DTP, DCT, NCI, Bethesda, MD 20205.
(8) Some 70 Material Classification codes exist. These are assigned to any active compound and range from 0D (active in P388, but insufficiently so to merit further testing) to 7A (commercial drug developed by NCI). The full list of Material Classification codes is published in "Instruction 14" (see ref 6).
(9) The first paper in this series contains some detail concerning the organization of the NIH Computer Center. Of note here is that the IBM 370 System and the DEC System 10 are "hard-wired" together. Passing large files from either of these computers to the other is very simple and can be very fast.
(10) Non-peak-hour transmission of data was used traditionally to minimize telephone charges. With the networking of the DCRT computers, which began in January 1986, the practice has become unnecessary, and it may transpire that true online operation will be feasible for the screening laboratories.
(11) It is relatively easy to convert a display-only field into a searchable field. The penalty that must be paid, however, is in the form of storage costs for the associated index files and overhead when files are updated. The decision as to whether a field should be searchable devolves therefore upon a trade-off of benefits vs. costs.

The NCI Drug Information System. 6. System Maintenance

M. T. ZEHNACKER, R. H. BRENNAN, and G. W. A. MILNE*

Information Technology Branch, Division of Cancer Treatment, National Cancer Institute, Bethesda, Maryland 20205

J. A. MILLER and M. J. HAMMEL

Fein-Marquart Associates, Baltimore, Maryland 21212

Received April 21, 1986

The NCI Drug Information System (DIS) is a collection of 24 interactively searchable databases which contain all the data associated with NCI's drug screening program. Data flow into all of these databases upon a daily basis, and maintenance procedures have been developed which provide a high degree of currency to the files. An extensive security system controls both write access and read access to the DIS and matches both to the authorization possessed by each specific user. Detailed usage statistics are collected automatically. The cost of the overall system in terms of both manpower and machine time is discussed briefly.

INTRODUCTION

The NCI drug screening program began in 1955 and is continuing at a pace that, although a reflection of fiscal limits, is still considerable. Currently, over 30 000 new structures are considered for screening each year, and between 5000 and 10 000 of these are actually acquired and tested. For any chemical that is acquired, a chemistry record is created at acquisition time, and then inevitably, records concerning its inventorying, shipments, and, finally, its biological test data begin to appear. In order that the NCI Drug Information System (DIS) databases be kept current, the system must be able to assimilate such data readily, and searches in the DIS must reflect the presence of the new data promptly. Given the flow of data and the number of databases, it follows that the updating programs must be essentially automatic, and this has been a design goal of the DIS.

Data integrity in the DIS has been supported, at least in part, by carefully restricting the number of individuals allowed to write to files in the system. To ensure this, a security system is used to control such write access to DIS databases. The same security system controls all read access to the DIS as well and, as will be described, monitors all use of the DIS.

An important guide to the continuing development effect in the DIS is found in the patterns of usage of the system. For this reason, detailed usage statistics are gathered by one of the management programs in the system, and these are processed into a report on a monthly basis. This report shows the frequency with which any command is used and also records the average cpu time required to honor the command. The statistics-gathering routine also records and reports any attempt to broach the DIS security on the part of any user of the computer itself, and such reports, when encountered, can be used to ascertain that data security was not compromised.

DATABASE UPDATING

A. General Considerations. When a new datum is acquired and must be added to one of the DIS databases, a number of steps are necessary. Modification of an existing entry is regarded as deletion of the old entry and addition of the new one, and so the same steps are involved. The first of these steps requires data entry, for which the DIS has an "update" program associated with every database. Thus CHEMUPD allows additions to the Chemistry database, NAMUPD to the Namecodes database, and so on.

With any of these programs, the user must first identify an old record that is to be changed or a new record which is to be entered. Once this is done, the new or changed data are

0095-2338/86/1626-0193$01.50/0

```
OPTION? NAMUPD

(A)DD, (C)HANGE, (D)ELETE, (I)NACTIVATE or (Q)UIT? C

NCOD to be Updated: 057B-3

Initiated By? BLANE
Source of New Information? MILNE

(INIT) Initiated By: BLANE
(SORC) Source of Information: MILNE
(NCOD) Name Code: 057B-3
(ADDR) Name and Address:
      Mr. Leonard H. Kedda
      DS&CB, DTP, DCT, NCI, NIH
      Landow Bldg., Room 5C-19
      Bethesda, MD 20892
(NAME) Name: Kedda,LH
(ORGN) Organization: NIH
(CLSC) Classification Code: 5;6;14;18
(GEOC) Geographic Code: E1119
(NCON) Normal Confidentiality Flag: 0
(SALU) Salutation: Mr. Smith

FIELD ? NAME
NAME is: Kedda,LH
Change to: Smith, A. B.
FIELD ? ADDR
ADDR is:
      Mr. Leonard H. Kedda
      DS&CB, DTP, DCT, NCI, NIH
      Landow Bldg., Room 5C-19
      Bethesda, MD 20892
Transfer Old data to OBNA  (Y/N) (Y) N
Change to:Mr. A. B. Smith,
      :DS&CB, DTP, DCT, NCI, NIH,
      :Landow Bldg., Room 5C-17
      :Bethesda, MD, 20892.

FIELD ?

Updated information for NCOD 057B-3

(INIT) Initiated By: BLANE
(SORC) Source of Information: MILNE
(NCOD) Name Code: 057B-3
(ADDR) Name and Address:
      Mr. A. B. Smith,
      DS&CB, DTP, DCT, NCI, NIH,
      Landow Bldg., Room 5C-17
      Bethesda, MD, 20892.
(NAME) Name: Smith,AB
          .
          .

NCOD 057B-3 has been Updated
```

Figure 1. Data entry step in an updating procedure.

entered interactively, field by field, until a new and acceptable record has been created. An example of this process is shown in Figure 1. Here, L. H. Kedda, having retired, is to be replaced in the Namecodes database by A. B. Smith. The approximate record, for namecode 057B, is first designated. Since this process will write in the database, the check-digit (the "3" in "057B-3") is mandatory. The program lists the full record for 057B and then asks who is initiating the change (Blane) and from whose information (Milne). The fields (NAME, ADDR) that are to be changed and the actual changes are then entered one by one until they are all made. Finally the updated record is printed and reviewed, and then the process is complete. With most databases, the new data move forward automatically to the index generation step, which is described below. An important exception to this is found in the Chemistry area. Because the accuracy of the chemistry data is so critical, any updated items in this database are treated just as new data and subjected to a special check by senior staff before being passed forward.

At this point, the displayable record has been edited, and any subsequent display of that record will reflect the changes. The searchable file, on the other hand, is created from the display file and does not, at this stage, reflect any of the changes. Consequently, a search for a newly entered value in a field will usually fail to retrieve that record because the pointer, or index file, that the search uses will still be in the form it was in prior to the entry of the new value. The way in which this problem is resolved is to discard the old pointer file after first having created a new one from the newly modified display file. During the time that elapses between entry of the new datum and generation of the new pointer file, the new datum will be displayable but cannot successfully be searched for. This is a short-lived and only mildly disagreeable state of affairs, because the displayable field is always accurate and the requirements of currency in searches in the DIS, while real, are less pressing than in, for example, a currency-exchange database or an airline-schedule database.

It is clear then that programs must be provided to regenerate the pointer files at regular intervals. Two practical problems exist, however. First, because searches will probably fail if they are attempted at the same time that the pointer files are being modified, some procedure must be developed to ensure that such contention for resources is not permitted. The means by which this has been achieved is described in the following section. The second problem is both more subtle and more significant. It concerns the precise strategy underlying the pointer-file regeneration and is taken up in detail in the third section.

B. Scheduling of Maintenance Work. The resource contention problem makes it necessary to schedule updating activities outside the normal working day, and even this, by itself, does not guarantee that the DIS will not be in use. Individuals can and do use the DIS around the clock, and some DIS use is deliberately scheduled for evening hours. Every night at 10:00 p.m. (Washington time), for example, all DIS data that was newly entered during the preceding 14 h is written to magnetic tape. This is done to guard against data loss as a result of system failure prior to completion of the necessary updates. In addition to this regular backup, offline use of the DIS often occurs during the evening, which is a good time to conduct very lengthy searches whose completion is nonurgent.

Because of the possibility of such activities, DIS maintenance work usually begins after midnight. Even then, DIS use—from people outside the Washington, DC, time zone, for example— is possible, and the updating programs therefore use a user-exclusion procedure. When the updating program begins, the first thing it does is to start a second program whose function is to access the database in question and, after checking that it is not being used, place a "lock" on it. The lock can prevent all programs except the updating program from reading a specific file or record, or writing in it, or both. If the database is found to be in use, the lock program retreats and returns after a short pause, typically 1 min, and will continue to cycle until it finds the database free. Once the database is locked, no other programs but the updating program can access it, and this will continue to be the case until the updating is complete and the lock is explicitly released. The details of the updating process are discussed in the following section. A typical update may take between 5 min and 1 h. The updating programs are all written as batch job control files, and they typically will contain numerous inner macros and FORTRAN programs. From the point of view of scheduling, the important outer controls concern the submit and the resubmit commands. The DEC System 10 batch controller allows one to submit jobs with a start time and a run time designated. A DIS update therefore will often be in the batch queue for hours or days awaiting its predetermined start time. Once started, the job will run until it finishes correctly or uses a predefined amount of cpu time. The cpu limiter is simply a guard against runaway jobs. When a job completes correctly, a full report of the job is printed, and the cost of the job is computed. The fact that it completed satisfactorily and the cost are passed back to an individual staff

```
From J. HOOVER          Postmark:21-Feb-86:00:01:29
To R. BRENNAN, J. WATERS

                    21-Feb-86  0:00:05

BATCON Version  104(4705)                GLXLIB Version  1(1136)

          Job SCAVAG Req #37 for JHOOVER [715,4466] in stream 4
 0:01:20 BATCH  .COST
 0:01:21 USER
 0:01:21 USER
 0:01:21 USER   Cost breakdown for J.HOOVER [715,4466]
 0:01:21 USER
 0:01:21 USER   February 21,1986 --Friday-- Time: 00:01:21
 0:01:21 USER
 0:01:21 USER
 0:01:21 USER   Category      Units      Charge   % Total
 0:01:21 USER
 0:01:21 USER   CPU time     5 seconds $   0.20     17.3%
 0:01:21 USER   I/O count      4489    $   0.64    54.0%
 0:01:21 USER   Core size      322P    $   0.34    28.8%
 0:01:21 USER                          +++++++++   ++++++
 0:01:21 USER   Processing cost to here $   1.20   100.0%
 0:01:21 USER   Connect time    00:01   $   0.02
 0:01:21 USER                          +++++++++
 0:01:21 USER   Session cost to here    $   1.22
 0:01:21 USER
 0:01:21 USER   **** Processing cost has been discounted by   60.0% *****
 0:01:21 USER
```

Figure 2. Cost report from a batch update job.

person via the electronic mail system that is resident on the computer and, finally, the job is resubmitted to run in just the same way at some date in the future, a week hence, for example. This then is a self-perpetuating process. Provided no errors occur, that particular updating process will start, run, and complete at regular intervals with no human intervention. All relevant new data entered since the last run will be swept together and merged into searchable files during the update, and all the staff sees is the completion and cost report, examples of which are shown in Figure 2. In this example, a job called SCAVAG[1] is reported as having begun at 5 s after midnight on Feb 21, 1986, and finished without error at 1 min 20 s past midnight, having used 5 s of cpu time. The cost of the job was \$1.22, having been discounted because it was run during off-peak hours.[2] The cpu time used by a job should be fairly constant from one run to the next; if it fluctuates sharply, the possibility exists that an error was encountered, and such jobs are examined further. Any errors are corrected, and if more cpu is truly required, the submit request is altered to reflect this.

The updating frequency of any DIS database is controlled by a single NCI manager. Some databases such as the Test Systems database change very slowly, and their updating is done on an as-needed basis.[3] Chemistry, Inventory, and Shipping data, in contrast, are generated on a continuous basis, and updates in these areas must be much more frequent. For practical purposes, a master update schedule is maintained and demands the major updates shown in Table I. As can be seen from the schedule, the work is concentrated around the weekends. The DIS is available for general use with no restrictions at midweek, but these update jobs may pre-empt users during the early morning hours between Friday and Monday. This schedule provides a satisfactory level of currency to the major DIS databases and, as will be discussed below, consumes perhaps \$6000 per month in computer costs. Higher updating frequencies could be undertaken, but this would drive up the cost and the present frequency is felt to be optimal in cost–benefit terms.

C. Details of the Updating Process. Updating of the inverted files in the DIS is much the same process as adding entries to a telephone directory. One way to do this would be to append the supplemental names to the existing directory and then re-sort the entire file so that the new names take their correct alphabetical place. Then the newly sorted telephone directory can be printed and distributed to users. If, as is usually the case, the supplement is much smaller than the old directory, this is rather inefficient because one has to re-sort a large file, most of which was sorted satisfactorily to begin with. A different approach would be to sort the supplement alone and create a true supplemental directory. The cost of this is minimal, and the real penalty devolves upon the users who would frequently have to look up numbers in more than one directory.

Table I

database	frequency	day
new data backup	daily	all weekdays
inventory	weekly	Tuesday
shipping	weekly	Friday
order	weekly	Saturday
pre-registry	weekly	Saturday, Sunday
letters	weekly	Saturday, Sunday
namecodes	weekly	Sunday
confidentiality	weekly	Sunday
chemistry	biweekly	Thursday
biology	biweekly	Tuesday

Likewise, in the case of a DIS database update, a supplement can be sorted and indexed independently of the main database, and the impact is felt primarily at search time because the search programs will have to consult both the main and the supplemental file or files. In practice, update supplements to the main databases are so small compared to the databases that the DIS almost always opts for incremental updating, in which the update increment is processed independently of the main file. As a matter of experience, over a period of 16 months, 32 separate incremental additions were made to the DIS Chemistry database. A search in the Chemistry file was, in reality, a series of 33 consecutive searches because of this, and it was also necessary for the DIS to merge the 33 results files so produced. All of this extra processing caused search times to increase, but by less than one order of magnitude; a search that should have required 5 s was requiring perhaps 40 s (elapsed times). At this point, the 32 increments, together with a 33rd, were consolidated together, leaving the main file and one relatively large increment. Searching was speeded up by a factor of about 8 with this consolidation, but a basal consolidation—a reinversion of the entire file—is still not justified, nor will it be for several years.

With the smaller databases in the DIS, consolidation is a more common event, although even here the general practice is to work toward a main database and a few increments. The number of increments is kept small by consolidation of increments with one another—but not with the main database. The single exception to this is with the Biology database. This is the largest file in the DIS, but it is overhauled once per year, as has been described in the previous paper in this series. An integral part of this overhaul is the reinversion of the entire file, once all the older data have been removed and archived. Because of this, regular updates to the searchable Biology file are processed as increments; no consolidation effort is made during the year between the major updates.

DATA SECURITY

Security is a significant concern for the NCI in operating the DIS. A major reason for this is that much of the data in the DIS concerns chemicals supplied to the government under confidentiality agreements.[4] These agreements stipulate that for those compounds the government will divulge no data to the public. In this way, the patent position of the owner of the compound is protected. The implication of this is that as far as users of the DIS are concerned, unless they are employees or contractors of the government, great care must be taken to ensure that they are unable to retrieve confidential data. A second, and minor, point is that database integrity

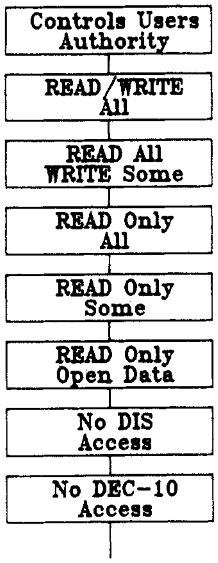

Figure 3. Hierarchy of user privileges in the DIS.

can be maximized if the number of individuals able to write in the databases is limited to as few as is feasible. Further, double-checking of all hand-entered data is advisable in order to eliminate as many errors as possible. These considerations have prompted the development within the DIS of a fairly stringent security system, which controls all reading and writing of DIS files and whose operation is described here.

Every compound and every sample of every compound in the DIS is classified as either "open" or "discreet".[5] These classification flags themselves constitute one of the DIS databases; it is a database which is closed to all but a very small number of people and which is updated upon a weekly basis.

Any user of the NIH computer may attempt to access the DIS, but access is permitted only to those whose identity is explicitly known to the DIS. Among those allowed to access the DIS, various levels of privilege are possible. One user may be authorized to view any DIS data, but another may only be allowed to see nonconfidential data. In between these extremes, a user who supplies compounds to NCI will be allowed to see any nonconfidential data as well as open or confidential data pertaining to compounds supplied by his organization. Further, while two users may both be allowed to view confidential chemical structures, it may well be that only one of these individuals is allowed to alter a chemical structure in the database.

When any person accesses the DIS, a DIS controller program notes their identity and reviews their privilege level. Every command subsequently issued by that person is weighed by the controller in the light of this privilege level. Only when it is clear that the user's authority will allow whatever activity is implicit in the command will the command be honored. If the command is a search that retrieves confidential data and the user is not authorized to view such data, the DIS detects this and removes all confidential data from the search results file before informing the user as to the completion of, and results from, the search. Every response the DIS makes to a user is first examined in terms of the user's authority. In this way, it is very easy to detect and prevent direct attempts to penetrate the database, and tests have shown that extremely indirect approaches, such as teasing out confidential data by means of complex Boolean strategies, are likewise precluded.

The overall structure of access privilege is shown in Figure 3. At the lowest level, access to the DEC System 10, as is normal, is restricted to those with valid accounts protected by passwords. This computer will not allow the use of short or guessable passwords such as initials and, furthermore, demands that users renew their passwords at regular intervals. Users

** DIS STATISTICS REPORT **
Feb-86

USER: A. B. Smith

COMMAND NAME	CPU TIME (AVG)	TOTAL COUNT
EXIT	2374	13
FORMAT	164	21
FIELDS	1053	1
TYPE	2526	29
XSEARC	1883	33
LISTQ	731	19
IDENT	1734	4
NUC	104	1
REG	435	2
SLGEN	203	1

FIELD SEARCHED, REGULAR	TIMES SEARCHED
HLDR	3
CNAM	4
NCOD	6
NAME	4
DACQ	4
FSUP	1
NSC	11
COMI	2
SSPL	2

OPERATIONS	TOTAL COUNT
FIELD SEARCHED, REGULAR	37
FIELD SEARCHED, EXISTENCE	0
SPECIAL REPORTS USED	0
FIELDS DISPLAYED	26
FIELDS SORTED	0
TOTAL NUMBER OF SESSIONS INITIATED	14
TOTAL NUMBER OF SESSIONS COMPLETED	10

Figure 4. DIS usage statistics.

of the computer may not, in general, access the DIS; the system maintains its own file of valid account numbers, and all other account numbers are refused access. All users of the DIS may read open data; to some users, no other privilege is allowed. Others may read some, but not all confidential data, and still others may read all confidential data. So far, however, writing has been precluded. The next two levels of authority allow partial and total write authority. The highest level in this scheme is that possessed by one person, the DIS System Administrator. In addition to all subsidiary privileges, this person alone has the authority to manipulate the User Privilege database, which is where the authority profiles of all other users are maintained.

This elaborate security system achieves three distinct goals. Most importantly, it rigorously guards against inadvertent release of confidential data. Second, it allows precise control over all writing in the databases and thus tracing and correction of error-producing situations. Finally, in the same way, it is used to control and record activities, such as orders to screeners, that imply government contractual authority.

USAGE STATISTICS

The standard DIS prompt is "Option?" and any response to this prompt is noted by the system and passed to a file of usage statistics. In this file is noted the precise command, the account number of the user, and the time, date, and cpu time used to deal with the command. At monthly intervals, this entire file is copied by a processing program which organizes the data by account number and prints a report similar to that shown in Figure 4.

In this example, the commands EXIT, FORMAT, etc. were each used the number of times noted during Feb 1986 by this user. The average cpu time (in milliseconds) required by the command is given, and it can be seen that cpu times on the order of 0.1–3 s are common here. The command XSEARCH denotes a standard alphanumeric string search, and 33 of these were recorded. All searches are counted in terms of the field involved and they are so listed, but their total (37) exceeds the XSEARCH count above. This is because a search statement such as

OPTION? CNAM/PHENYL AND DACQ/1985

is counted as a single XSEARCH, but two fields are involved and are both counted once. As is common, the NSC Number field (NSC) is prominent in this list. Finally, a summary for this user is presented. The total number of field searches completed was 37, which is also the total of the "times searched" column. The number of different fields displayed was 26, and the user accessed the DIS 14 times during the month, leaving the system by means of an "exit" command 10 times and 4 times by some unspecified route.

COSTS

The Drug Information System is comprehensive in that it supports the entire drug screening effort at NCI, encompassing selection, acquisition, shipping, storage, and testing of chemicals for anticancer activity. The decision to develop such a computer system was taken with deliberate care because it represented a significant investment on the part of the Institute. Development of the DIS was viewed as inescapable, however, because the needs of the program were very pressing and no commercial database management system, then or now, offered any hope of meeting the needs that had been identified.

Design, development, and implementation of the DIS began in March 1982, and the bulk of the whole system was installed during the latter half of 1985. At the end of that calender year, the single contractor had expended approximately 20.1 man-years[6] on the DIS, and the cost of this labor was $1.3 million. The computer costs incurred during the same time were on the order of $2 million,[7] and other costs of $300 000 had also been incurred.[8] Thus, the DIS, in its first essentially complete form, cost NCI about $3.6 million. This, it might be noted, supports the NCI drug screening program that currently operates at an annual level of $30 million. The probability of discovering medically useful drugs is low, but such compounds are produced by the program. Of the 30 cytotoxic drugs commercially marketed in the U.S., approximately 14 were discovered and developed by the NCI, which has also made major contributions to the development of most of the other 16.

Since the initial installation, the NCI has supported a maintenance contract whose annual cost is now just under $150 000. Two full-time positions are provided under this contract and are able to deal with both maintenance of and enhancements to the DIS. Within the NCI, three persons each spend perhaps 40% of their time updating the DIS databases. The monthly computer billing for this updating is currently on the order of $6000, and so the overall ongoing cost of the DIS is approximately $220 000, not including NCI staff effort. The cost of the data entry is omitted from this summary because it is not easy to estimate with accuracy. Three people enter data on a full-time basis, and as many as 40 people do so as a part of their responsibilities. It is probably reasonable therefore to assume that data entry accounts for the equivalent of perhaps seven full-time people, and the cost of data entry is therefore on the order of seven middle-level salaries.

Design, development, and installation of a system like the DIS costs therefore some $3 million, and the continuing support of the system is something under $400 000 per year, i.e., 13% of the start up cost or 1.3% of the overall program cost. As will be perceived by readers of these papers, the DIS is large and complex; these features, judged to be necessary by NCI users, clearly underlie much of the cost. Database management systems of many sorts are available commercially, but none were found, then or now, that could approach the level of support offered to NCI by the DIS. Indeed, the DIS is so custom-tailored to the peculiar needs of the NCI that one should not expect otherwise. It certainly is true that for far less money any of the common database management systems could have been installed in NCI. As acquired, however, it would not have met program needs, and adequate modification to meet those needs would have implied effort and costs tantamount to what was in fact spent on the DIS itself.

REFERENCES AND NOTES

(1) SCAVAG is a housekeeping program which collects all shipping requests that involve NCI's Brussels office. It is run once per week, and all the shipping lists it finds are printed and sent to Brussels. As can be seen from the log, SCAVAG is a trivial program in terms of run time and cost.
(2) Outside the normal office hours (8:00 a.m.–6:00 p.m.) the cost of using the NIH computers decreases by as much as 60%.
(3) New test systems are added to the screening protocols as they are developed and deemed useful. Such new test systems are adopted a few times per year at most.
(4) Confidentiality is a very important condition of the agreements under which companies supply the government with free samples of chemicals. The so-called "discreet agreement" by itself represents a legally binding contract, which NCI observes meticulously.
(5) Confidential compounds in the DIS are described as "discreet", and their NSC Numbers are prefixed with a "D". The misuse of this word has long been recognized; it is a practice, however, which is so ingrained that no amount of reference to dictionaries can be expected to change it.
(6) The figure of 20.1 man-years covers the contribution from the main DIS contractor (Fein-Marquart Associates of Baltimore). Various other organizations have contributed to DIS design and development as a part of their normal support role to NCI. These contributions are not counted here, nor is the work in systems development carried out by NCI staff.
(7) The figure of $2 million covers all work done on the DEC-10 by the Information Technology Branch during the period in question. Most, but not all, of the work was DIS related, and so the "DIS cost" of $2 million represents a high estimate.
(8) These other costs relate to purchases and lease of hardware, including computer terminals, and a variety of other minor expenses, such as consultant's fees and support staff within the government.