

SDI System for the U. S. Public Health Service, Office of Pesticides*

MOLLY A. WOLFE and SAUL HERNER
Herner and Company, 2431 K Street, N. W., Washington, D. C. 20037

Received May 10, 1967

The SDI system described was designed to relate the activities and information needs (determined by interviews) of users with the pesticide literature. The information is described by a classification system derived from an analysis of the mission of the Office of Pesticides and the literature relating to the mission.

In May of 1966 we undertook to design and operate for The Office of Pesticides of the U.S. Public Health Service a retrieval system, a monthly abstracting publication, and a selective dissemination system (SDI), all dealing with the health and medical aspects of pesticides. This paper deals with the SDI portion of the project.

The audience for the pilot SDI service consisted of professional members of the staffs of the Office of Pesticides, other Government agencies with a concern for the effects of pesticides, such as the Food and Drug Administration, the Communicable Disease Center, the Federal Committee on Pest Control, the National Cancer Institute, various units within the Department of Interior, and outside contractors involved in the community studies program of the Office of Pesticides.

The specific mission of the SDI portion of the project was to produce monthly notifications for the test audience. These notifications were to contain full bibliographic citations and informative abstracts. The dissemination of notifications was based upon a matching of a hierarchical subject classification of the contents of the documents against an identical classification of the activities and work-related interests of the recipients. The classification scheme was such that matches between the activity and interest profiles of the recipient and the contents of documents could be made at either the class or subclass level or by means of coordination of classes and/or subclasses, thus allowing matching flexibility and resolution.

DEFINITION OF SCOPE

The first step in the development of the SDI service was to define the subject scope of the documents to be collected and analyzed. This was done in a number of phases, the first involving the analysis of the exact mission of the Office of Pesticides, and the subjects and activities implied in this mission. This preliminary assessment was followed by the preparation of a tentative list of topics which was presented to the Office of Pesticides and refined over several successive stages. The listing of topics with which we began our SDI service was ultimately the prod-

uct of many changes, and it has changed further in the course of our operation of the program. There will obviously be many other changes as the field evolves and a fuller understanding of its implications develops.

The definition of subject scope was followed by a detailed analysis of potential sources of information on the topics defined. This was done *via* a broad-based scanning of the literature holdings of several major source libraries in biomedicine and agriculture, through analyses of all available bibliographies dealing with the topics and subtopics defined, and through statistical definition of those publications having the highest probability of producing the information sought. As it developed, the field under consideration is of a highly multidisciplinary nature, and the sources of information that have to be covered are extremely diffuse, consisting of several thousand biomedical, agricultural, and chemical periodicals and serials.

USER STUDY

Another means of determining probable sources of inputs to the system was a detailed interview study of its prospective users. The purpose of the user study was to define the various functions that the total system, of which the SDI system was a part, was to perform. The parts of the interview survey that were particularly germane to the SDI system dealt with the following: the professional backgrounds and activities of the users, their subject training and interests, their professional society memberships, professional meetings attended in the previous two years, titles of their publications, their language capabilities, titles of journals and other publications seen regularly, abstracting and indexing publications seen in the previous year, words and phrases in publications scanned that would prompt detailed reading, review publications seen regularly, the subject of bibliographies that they prepared or had prepared for them, the subject and language of translations done or prepared for them, and the subjects covered by any group or personal files maintained in connection with their work. The answers to the questions were examined in detail in developing not only the subject profiles of the prospective users of the system, but also the probable sources of information that would be most responsive to their needs.

* Presented before the Division of Chemical Literature, Symposium on Selective Dissemination of Information, 153rd National Meeting of the American Chemical Society, Miami, Fla., April 10, 1967. Work performed under Contract PH-86-66-133, with the Office of Pesticides, Communicable Disease Center, U. S. Public Health Service.

DESIGN OF PERSONNEL AND DOCUMENT
PROFILING PROCEDURES

As noted, the profiling of both the prospective recipients of the service and the documents was based upon a hierarchical subject classification which permitted matches between activities and interests and the contents of documents at either the class or subclass level or by means of coordination of classes and/or subclasses. The reason for the use of classes and subclasses was twofold: First, as indicated from results of this and prior user studies, people generally think in terms of broad concepts rather than specifics in seeking and attaining their current awareness requirements; second, we were interested in providing for the greatest possible flexibility and discrimination in matches between interests and the contents of documents.

After detailed analysis of user requirements and the most responsive approaches to current information within our prospective audience, we ended up with three basic classes or categories of information. One was a broad or generic class which dealt with the various aspects of the basic mission of the Office of Pesticides. This included the following topics: research and monitoring, poisoning cases and their treatment, research in toxicology and pharmacology, safety, methods of analysis, and general topics such as news and editorials and regulations and legislation. The second class or category dealt with classes of chemical compounds prominently used as pesticides. The third class or category dealt with the biological and agricultural aspects of the subject, such as types of organisms, test systems, methods of application of pesticides, and the use of different classes of pesticides.

INPUT PROCESSES

As shown in Figure 1, initial input for the system consisted of two four-step processes. One process, which is performed at the onset of the system operation and is refined and augmented on the basis of regular feedback or responses from the recipients, is the conversion of the responses to the key questions in the interviews into classes and/or subclasses, which then constitute interest profiles. The second process, which is performed on a monthly basis, involves the assignment of classes and/or subclasses to incoming documents.

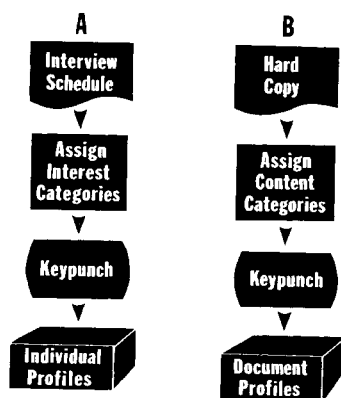


Figure 1. Input processes: A, document profiles; B, individual profiles.

A second major part of the input process is given to the preparation of informative abstracts, the transcription of the abstracts plus bibliographic citations onto punched paper tape, and transcription to magnetic tape for subsequent computer manipulation (Figure 2).

MATCHING PROCESS

Figure 3 gives an over-all view of the processes involved in the monthly matching of individual and document

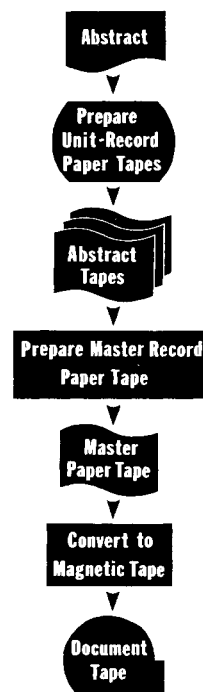


Figure 2. Preparation of document tape.

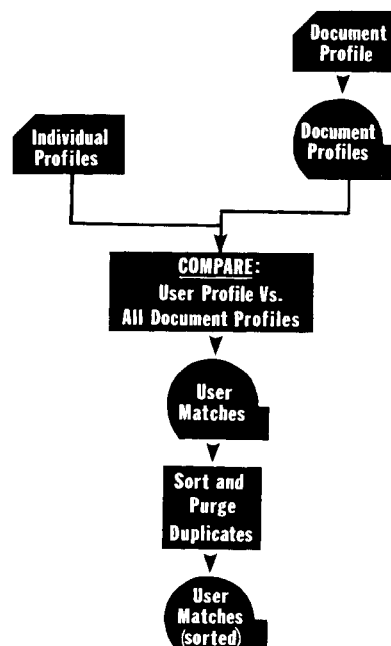


Figure 3. Matching processes.

Input: Document Profile Cards
Individual Profile Cards
Output: User Match Tape

profiles. Each month, the subject profiles of all incoming documents and the profiles of the individual recipients are read into an IBM 1401 computer *via* punched cards. The individual profiles are matched against the document profiles for coincidence of classes, subclasses, or prescribed coordinations among them.

Figure 4 illustrates four matches between coded individual and document profiles. In the first match, the person is interested in anything on the toxicology and pharmacology of pesticides. This is an entire class. The documents produced by this match deal with acute toxicity, chronic toxicity, cholinesterase and other enzymatic inhibition studies, metabolism, persistence-tolerance mechanisms, interactions, and therapy evaluation. In the second match, the person is interested only in the subclass dealing with metabolism. In the third match, the person is interested in the metabolism of organothiophosphoric acids and their esters and salts, which represents a coordination between two subclasses. In the fourth match, the person is interested in the metabolism of organophosphorus compounds in man, which represents a coordination of a subclass (metabolism) with a whole class (organophosphorus compounds) and another subclass (man).

Document Numbers	Document Profiles		
1	AC01	CC03	
2	AC04		
3	AC04	CB04	
4	AC04	CC02	
5	AC02		
6	AC04	CC03	RA01
7	AC05	CC02	RA01
8	AC04	CC02	RA02
9	AC06		
10	AC04	CC02	RA01

Individual Profile	Matching Documents
AC	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
AC04	2, 3, 4, 6, 8, 10
AC04 CC02	4, 8, 10
AC04 CC RA01	6, 10

Figure 4. Matches between individual and document profiles.

OUTPUT PROCESSES

Figure 5 is an overview of the output processes of the system. A tape containing recipient numbers and document numbers representing matches is compared against a document tape containing document numbers, citations, and abstracts. Where a coincidence between a recipient-document match and a document number on the document tape occurs, the source citation and abstract for the document are printed and forwarded to the recipient along with a response card that is used for quality control purposes.

Another part of the output process involves the cumulation and printing of document and user statistics. The statistics maintained give the number of documents

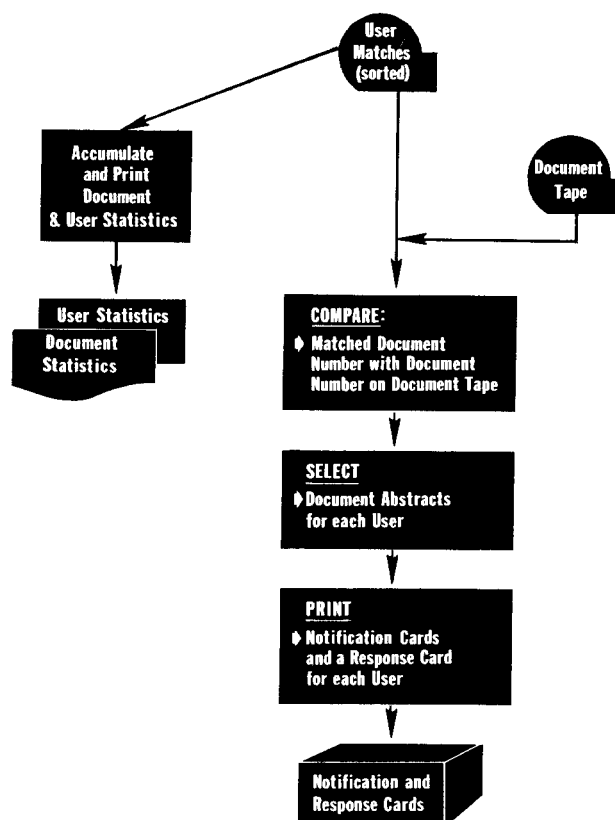


Figure 5. Output process.

INPUT: User Match Tape (sorted) Document Tape

OUTPUT: User Statistics Document Statistics
Notification and Response Cards

matched against each user and the number of users who receive each document in a given month. This serves as a presumptive test of the degree of sensitivity of resolution obtained by the system.

QUALITY CONTROL

In addition to the monthly document and user statistics, the system uses two other mechanisms for measuring the quality of its performance. The first, already mentioned, is the monthly user-response card on which the recipient indicates which of the notifications were pertinent to his interests and which were not. The second quality control mechanism is a questionnaire in which the recipient is asked a series of questions regarding his mode and degree of use of the notifications sent him, problems and indicated improvements, and significant omissions in the notifications.

The analysis of omissions is particularly important, because it provides, along with the monthly user-response cards, a fairly rigorous means of evaluating what the system is and is not doing. The specific means by which significant omissions are determined is to have the recipient match his notifications against the contents of a monthly *Literature Bulletin* which encompasses all of the inputs to the system. This provides the means of detecting oversights, and serves as a guide for the refinement of the user-profiles and matching logics to make them more completely responsive to the recipient's needs.

On the basis of the operation of the system during its first three months, the percentage of nonpertinent documents averages approximately 26%, and the percentage of pertinent documents missed averages 5%. Our continuing quality control efforts will presumably improve these figures, particularly that for nonpertinent documents.

In one series of questions in the questionnaire, the respondents were asked whether and why they preferred the SDI notifications or the *Literature Bulletin*. (All recipients of notifications also receive the *Bulletin*.) Half stated no preference or specifically stated that they felt both were necessary; 15% preferred the *Bulletin*; and 35% preferred the notifications. Among those who preferred the *Bulletin*, the primary reasons were because of the availability of a table of contents, broader coverage, indexes, and better readability. The prime reasons for preference of the notifications were that they provided a means for setting up personal files, and they provided a selectivity which was not available in the *Bulletin*.

TIME AND COST ASPECTS

Regarding the time and cost aspects of the system design and operation, the system was fully operational and the first notifications were sent out within five months after the onset of the project. The design consumed approximately three months of the time of a senior systems analyst and about 50 hours of programming time. In addition to the design aspects, the time of the senior analyst was given to the development and polishing of profiles for the members of the pilot audience, which consisted of 100 persons.

The present input to the system consists of approximately 140 items a month. This produces an average of about 40 notifications per person per month. The average notification appears within eight weeks following the issuance of the original publication. The seemingly high number of notifications per recipient is due to the specialized nature of the literature covered and the commonality of the activities and interests of the audience.

The monthly costs of the operation of the system, stated in terms of man-hour, machine, and supply expenditures, are approximately: $1\frac{1}{4}$ the time of an abstractor, $\frac{1}{7}$ the time of an indexer, $\frac{1}{2}$ the time of a Flexowriter operator, 2 hours of a keypuncher's time, 1 hour of converter time for paper tape to magnetic tape, 6 to 7 hours of IBM 1401 computer and operator time, 4 hours of clerical time for bursting cards, stuffing, mailing, etc., and about \$200 a month for supplies such as Flexowriter tapes, magnetic tapes, cards, etc.

CONCLUSIONS

It is, of course, dangerous to attempt to draw excessively expansive conclusions about an experimental system that

has been in operation for less than one year. However, certain tentative conclusions seem indicated, if only as bases for further examination and validation by us and others.

First, in regard to the system concept and design, when we started this undertaking we had a choice of attempting to adapt a "canned" SDI system and program or developing our own. We chose the second course, first, because most of the available systems were based on matches of words rather than the classes and subclasses which we felt to be essential for matching flexibility and resolution; and second, because preliminary systems analyses indicated that the costs and complexities involved in adapting and operating the "canned" systems far exceeded those involved in designing and operating our own, as we had envisioned it. In essence, we tried to substitute conceptualization for computerization, and to keep the system as simple and transferable as possible.

The basic reason for our urgent desire for simplicity, aside from the need to transfer the system to our client with a minimum of difficulty, was that we harbored, and still do, some doubts as to the place of SDI systems in research programs, and we did not want to overcommit prematurely in this direction.

When one thinks in terms of a huge and diffuse literature such as that covered by *Chemical Abstracts*, it is easy to see a distinct advantage, in time and timeliness, in a system that sifts all available sources of information and directs to the user only those items that are germane to his activities and interests. Such a system also has clear advantages in mission-oriented situations, where the literatures of a multiplicity of fields may be germane but difficult to cover.

However, from our evaluative survey, there is tentative evidence that SDI is not universally applicable or acceptable in all research situations, particularly in relatively narrow fields in which specialized abstracting and indexing publications are available, as in the present case. In our survey we found that the prime users and appreciators of the SDI service were persons working in the periphery of the field, and those involved in the clinical and other applied aspects of it, while the smallest users and appreciators of the services were the pure scientists, who generally preferred to make their own selections from the *Bulletin*. This follows quite logically on the results of past user studies, in which it has been shown that pure scientists are generally reluctant to delegate their information-seeking tasks, while persons involved in applied work are more prone to delegate.

Thus, we are not only concerned with evaluating and refining the system we have developed, but also with examining and identifying the place of the SDI concept in the spectrum of information tools and services. It is always important to avoid treating new and interesting tools and techniques as panaceas, and thus assuming the risk of having them discredited because of misapplication.