

November 13, 1967

PROGRESS REPORT ON THE IS&R SYSTEM

This is a progress report on the development and operation of the IS&R System. The report sets forth what has been accomplished to date, and some recommendations for the future. It was prepared at the request of Miss Brown, acting on behalf of Mr. Hetsko.

I. Current Status

The system is in operation and has the capability to provide rapid scientific documentation in response to user inquiries, in accord with the original objectives set forth by users. These objectives, as we understand them, were to:

- 1) Establish a mechanized system that is not dependent on the knowledge, experience or availability of a few individuals.
- 2) Provide for fast acquisition from the system of selected materials for specific congressional, public relations and other purposes.
- 3) Provide for quick answers to specific questions on the pros and cons of Tobacco and Health.
- 4) Provide for regular distribution of user - oriented abstracts of pertinent new research papers.
- 5) Enable the users to meet the efforts of critics, who have up-graded their research analysis and collection procedures, have a computer and considerable manpower available.

A. Materials in the System

There are now approximately 10,000 documents stored in the computer, each of which has been abstracted and indexed to an average depth of 50 terms. The materials now stored in the system include:

- 1) All relevant materials published after January 1, 1963 that are known to be available to the National Clearinghouse for Smoking and Health (NCSH). These include references published in the three major bibliographies issued by NCSH, as well as the monthly NCSH bibliographies.
- 2) All relevant papers published since December 1966 when the IS&R system became operational. These papers were selected in regular 3i scanning of some 3,600 scientific and medical journals.
- 3) Selected, highly relevant papers published prior to 1963, as provided by various users and the project officer. This is a limited number.
- 4) Selected unpublished materials as provided by various users and the project officer (e.g. key papers presented at meetings, unpublished articles by such authors as Hammond, Auerbach).

3

Documents are being entered into the system at the rate of about 1,000 a month. All materials entered into the computer are selected on the basis of guidelines developed by a committee of users. (The Scope of Coverage for the system is found on page B-1 of the User Manual, distributed to all users in June 1967.

B. Current Capabilities of the System

The system provides rapid access to the relevant scientific and biomedical literature published after January 1, 1963. The key to access is in-depth indexing.

Documents can now be retrieved by any one of the following characteristics, or any combination of these:

- . Author's name
- . Author's affiliation
- . Subject category
- . Year of Publication
- . Journal of publication, or place of delivery, in the case of unpublished work
- . Index terms, or descriptors -- averaging 50 terms per document.

Also, through the precoordination of terms, users can retrieve documents on the basis of the types of opinions expressed, concessions made, and helpful information provided -- either by specific authors, or on general or specific subjects.

For example, it is possible to retrieve all concessions by Hammond regarding statistical association. Or, to retrieve all papers in which the authors concluded that smoking is not known to play a significant role in causing lung cancer, or other diseases. It is also possible, for example, to rapidly retrieve all papers on smoking and lung cancer in which the authors express the opinion that smoking causes the disease, but also acknowledge that their opinions have not been proved.

Because of the depth of indexing, and the flexibility of the computer programming, this system provides retrieval in greater depth, and with greater specificity than any other IS&R system known to be operating.

II. Working Tools Provided to Help the User

- A. Thesaurus -- The thesaurus is being distributed to all users this week. Its use should increase understanding of the system and efficiency in using the system.

Much like a classified telephone directory, the thesaurus is an alphabetical cross reference guide to specific and general subjects stored in the system. It provides a compilation of the descriptors in the computer and shows cross references and interrelationships between the various descriptors.

The thesaurus represents the third stage in the development of tools that will enable the user to make searches more effectively. Earlier steps, each necessary prior to development of the thesaurus, were:

- 1) The Permuted Authority List (or KWIC List) -- This is a complete listing of searchable descriptors arranged alphabetically by consti-



tuent terms. Each descriptor listed is preceded by a number showing the frequency with which this descriptor had been used, at time of issuance.

- 2) Three Updated Authority Lists -- These are alphabetical lists of descriptors stored in the computer. Further updated Authority Lists will be issued at intervals, as new data are entered into the computer.

B. User Manual -- The first edition of the User Manual was distributed to all users in June 1967. The manual describes the system and its capabilities, tells how to form a search question, how to fill out the search request form, and how to communicate with the system. It is indexed in depth for easy reference.

The section on "How to Communicate With the System" outlines the alternative methods of getting information from the computer. You can, for example, ask the project officer (or 3i in his absence) to formulate your questions, without learning how to phrase search questions yourself. Or, you can learn to form your own questions, and then communicate through the project officer, or go directly to the computer location. These alternatives are available because it is recognized that not all users will wish to learn how to "work" the system, while some will prefer to become experts on the system, its contents and its capabilities.

In short, one need not be an expert to use the system. Available experts can do the question phrasing.

In addition to these working tools, the project officer has held training sessions with some individual users. Representatives of 3i are prepared to take part in further training, if necessary, in cooperation with the project officer.

III. Informing the User of the Important New Articles

Through regular scanning of 3,600 journals, 3i is continually uncovering published research that may be important for users to know about quickly. Alerting users to this information is, of course, an objective of the system.

To meet this objective, the Current Awareness Bulletin has been distributed bimonthly by the project officer to all users.

However, experience has suggested that the present Bulletin provides too much material for some users, perhaps not enough for others.

Because of the diversity of interests among individual users, 3i has proposed a more refined current awareness system. Through Selective Dissemination of Information, each user would receive information tailored to his individual needs and interests. Those who want extensive coverage of particular subject areas can obtain it. Those who want minimal coverage can get that. Those who want minimal coverage of 15 subject areas, for example, and extensive coverage of three areas can get that.

The project officer and 3i plan to contact each user to develop a profile of interests for each individual user. This profile will form the basis for the new current awareness system. (A copy of this proposal is attached.)



IV. Use of the System

The attached report from System Science Corporation, 3i's computer subcontractor, gives details on usage of computer time to date. As this report shows, the reserved computer time has not been fully used. However, since some users have not yet begun to use the system -- due to lack of familiarity with it, or the pressure of other work -- it is probably premature to conclude that too much time has been reserved.

The project officer advises that the system has been a helpful tool to those who have actually used it. (We understand the project officer has been asked to prepare a progress report, which presumably will discuss this question.)

The system has also been effectively used by 3i personnel in connection with special Ad Hoc projects (e.g., search out papers showing nonassociations and negative associations between smoking and various diseases).

Beyond this, there is little 3i can state from its own experience regarding use of the system. This is because 3i does not at present see the questions asked by users, nor the computer's answers. The users and the project officer may wish to consider whether 3i's experience in preparing the input and in developing the overall system could be more effectively utilized in helping various users become proficient in using the system, and in developing refinements of the system.

V. Proposals Which Can Be Implemented Within the Present Budget

- A. Increased regular communication (feedback) between users and the system operators (3i and the project officer) is important.

Originally, when we first discussed development of the system, 3i emphasized that continuing communication and user feedback were important ingredients in refining both the input and the output of the system. The project officer provides some of this to 3i. However, feedback to him appears to be somewhat limited, and he is limited in what he can provide 3i by virtue of not being a prime user.

- B. Early evaluation of the operation of the system by a user committee is needed.

Here are some of the questions that should be considered:

- 1) Is the scope of coverage adequate and meeting user needs? Or is it more extensive than necessary?

It is 3i's view that the scope of coverage could be reduced somewhat. It had been our understanding that there would be a review of this matter by a user committee; this was apparently postponed because of other pressures. This step should be taken soon, however, while the system is still in the early, developing stage. It can be done in great part by user feedback, as well as through a meeting.

- 2) Is the indexing effectively meeting user needs? Should it be refined or revised? Does user experience demonstrate the need for continuing to index to an average depth of 50 terms per article? Or, would fewer terms fill the need as well?



- C. Prepare in advance for future needs. Develop the answers to certain fundamental questions, or points; store these questions and answers in the computer; regularly update them.

This idea would save some time and money. Moreover, it would provide each user with ready documentation to support certain basic points (e.g., statistics do not prove cause and effect; multiple factors are involved in lung cancer and heart disease; nonsmokers get these diseases). This idea was originally contemplated by some users, and has been renewed by Miss Brown.

- D. Enter the 1965 and 1967 cigarette hearings into the system, as well as the HEW Appropriations Hearings for 1968.

We understand these hearings contain a good deal of helpful information and concessions. It might not be necessary to enter the full hearings. Rather, the volumes could be indexed chiefly for the helpful material, with this material entered in the computer.

- E. Enter selected pre-1963 materials into the system.

- 1) Some 2,000 pre-1963 references in the latest bibliography from the National Clearinghouse for Smoking and Health.
- 2) Selected pre-1963 papers from among the collection of the Council for Tobacco Research. Selection of 1963 papers was originally deferred until the system was operating, and 1963-67 papers had been entered. Selection of papers to be entered could be determined following establishment of guidelines by the users.

- F. Enter into the system legal and public relations work products developed by individual users and other industry personnel.

This was an area of particular interest to some users during early planning discussions. The techniques and methods for entering work products can readily be determined, including those to insure the necessary confidentiality. Some examples of the type of work product that might be included are: rebuttals to particular theories, such as selenium, nicotine and "tar"; and rebuttal to particular authors, such as Auerbach and Hammond.

- G. Experimental use of the system could be encouraged to refine question-asking techniques, to check the adequacy of indexing and scope of coverage, and to increase familiarity with the system.

VI. Other Proposals (Involving Some Added Cost)

Incorporate the thesaurus into the computer. This step would enable the computer to do most of the question formulation. All cross-references, narrower terms, related terms, etc., would automatically be searched by the computer when a user asked a question. For example, if a user searched the term LUNG CANCER, the computer would automatically search such terms as PULMONARY CARCINOMA, PULMONARY ADENOCARCINOMA, PULMONARY BASAL CELL CARCINOMA, etc.

3

This step could be taken following some experience in using the new thesaurus, and allowing time for refinement and updating of the thesaurus.

VII. Miscellaneous

Miss Brown asked specifically about any present activities related to the IS&R System that had to be conducted manually because of the confidentiality requirements.

One requirement of this nature had resulted in manual operation in some instances. This is the requirement to limit computer input in such a way that bibliographic print-outs are not possible.

- When 3i was asked to help evaluate three Public Health Service bibliographies, for example, the first step was to determine which of the PHS references were in the system. This had to be done manually, and involved a good deal of time.

It would be possible to enter bibliographic data and maintain confidentiality. For example, bibliographic data could be entered, with output limited to accession numbers -- as at present -- except where bibliographic output was essential and authorized.

