

SIMON O'SHEA
919 18TH STREET N. W.
WASHINGTON, D. C. 20006

737-9038

MEMORANDUM

TO: Covington & Burling SUBJECT: Report and Review of IS&R System
FROM: Simon O'Shea DATE: 16 November 1967

Eleven months have elapsed since the decision was taken develop a Legal Information Storage and Retrieval System. Today such a system is operational. This memorandum traces some of the events in the history of the System's development and advances suggestions to aid in its evaluation at this 11 month juncture.

The decision to develop a computerized information system grew out of a number of factors. The volume of bio-medical, tobacco and health literature had been growing so steadily that it was becoming difficult to keep track of such a large amount of data with traditional, manual library techniques. As a result, scientific information of recognized value was being forgotten and eventually lost in a system that depended on human memory for its retrieval capabilities. Furthermore, speed of retrieval had come to be recognized as an important need when situations arose in which the industry was confronted with bibliographies running to 500 or more citations or when witnesses during an hours testimony cited dozens of scientific papers.

Developmental Period

Once the decision had been taken to create an Information Storage and Retrieval System, the actual developmental period began. During this period, 3i, in conjunction with various committees and the Project Officer, actually defined the specifications of the System that had been rather generally described in the original proposal.

This involved the drafting of instructions to 3i setting forth precisely what documents from the bio-medical literature should be selected as input for the Information Storage and Retrieval System. These instructions are called "The Scope of Coverage". Though this Scope of Coverage is a provisional outline which undoubtedly can be improved in the light of experience. It has, nonetheless, provided the criteria necessary to 3i for the operation of the System.

In addition, during the developmental period, instructions were prepared in consultation with the users setting forth the manner in which each selected scientific document would be summarized and indexed for input into the System. These instructions attempted to anticipate the kinds of questions that would ultimately

be asked of the System by the users, and to insure that each document was read, annotated, and indexed with a view to retrieving information of prime importance to the users.

These instructions appear in various memoranda defining: descriptors; descriptor components; levels of descriptors; the nature of annotations; conventions for citation information; and checklists of information to be recorded by the indexers. These first two steps primarily involved the literature itself and various documentation techniques.

At the same time, developmental work went forward on the computer programming. Here it was necessary to develop a computerized system that could absorb an enormous amount of data and provide flexible and rapid answers to user's questions.

Finally, it was necessary to develop and supply to the users all of the supporting artifacts of the System, such as: the aperture cards; authority lists; indexes; the specifications of storage, viewing and reproducing equipment; and instructional materials to aid the user in the operation of the System.

Operational Period

By early June the developmental period was largely completed and the capabilities of the System were made available to the users. Since June 1st the System has been operational and a summary of its use is set forth in Attachment A. Over the period of the summer and early fall, briefing sessions were held with virtually every user. During these sessions all of the elements of the System were reviewed and the users were encouraged to actually formulate search inquiries which were then transmitted to the computer.

During the operational period (June to the present) the System has proved of value, particularly in connection with various analyses of the bibliography of "The Health Consequences of Smoking" and in connection with preparatory work for the Congressional Hearings which ultimately did not take place.

Utilization of the System

Some idea of the kinds of questions currently being addressed to the System can be seen by referring to Attachment B. In general, most users have reported locating valuable, relevant data as a result of IS&R System searches. On the other hand, though the data base is now only about 10,000 documents, a few users have commented that search results are providing more drops than they can conveniently analyze.

So far no real pattern of utilization can be said to have emerged, although there seems to be a long term trend towards increased utilization. The October utilization of the System was inflated by a special ad hoc project carried out by 3i and by an experimental study of the effects of smoking on various parameters of blood chemistry.

Full utilization of all of the computer time available to us on a monthly basis (700 minutes) is certainly desirable. However, it would be a mistake to try to judge the value of the System simply on an arbitrary 700 minute scale. This 700 minute budget was predicated on the assumption that roughly 350 questions per month would be asked of the System.

Did the users, prior to the development of the System, frequently ask 350 literature research questions per month? In order to reach this level of utilization, every individual user would have to formulate at least 26 inquiries per month.

In fact, many users have been so busy with ad hoc, legal, and public relations responsibilities that they have only recently been able to secure the supporting equipment they need in their own user locations in order to use the System effectively.

Current Status

The fact is that we have fulfilled many of the goals set for the Legal Information Storage and Retrieval System 11 months ago. We have now absorbed over 10,000 documents from the bio-medical literature into the System. These documents can now be rapidly searched by a number of characteristics, most importantly by author and by thousands of subject descriptors.

The trained personnel involved in the System are now reading approximately 3,500 bio-medical journals a month, selecting material which we believe will be of utility to us, and annotating and indexing this material so that its informative content is rapidly available to the users.

Evaluation of the System

Since early June (the beginning of the operational period) most users have had some exposure and gained some familiarity with the System. With this background it should be possible to begin to analyze the utility of the Information Storage and Retrieval System as it exists now, with a view to recognizing what has been accomplished, as well as what may be modified or improved. Such a review might well include some or all of the following questions:

1. Are the documents the users need actually being selected for input? This question encompasses both the possibility that useful material may have been overlooked or that peripheral materials are being selected. An answer to this question would involve reviewing a selection of recent accessions, as well as the Scope of Coverage.

2. Are there additional materials which might usefully be proposed for inclusion in the System, such as a selection of pre-1963 bio-medical literature; indexed transcripts of relevant Congressional Hearings; various legal materials;

prepared subject bibliographies; public relations background memoranda; and other materials.

3. Are the new documents being added to the System actually being annotated and indexed in a manner that insures their utility for the users? Here again a user review of recent accessions (i. e., aperture cards) would be helpful.

4. Are computer searches working? Are answers useful and provided in a timely fashion? Could search strategy be improved?

5. Are the System's supporting materials, such as aperture cards, authority lists, The Current Awareness Bulletin, the Thesaurus, and others adequate for convenient use?

6. Are there improvements to the System that should be planned for the future? There are a number of such improvements which have been suggested, including:

a. Programming the Thesaurus into the computer so that there would be an automatic cross-reference feature to searches.

b. The development of an author index which would list all of the authors stored in the computer, along with the accession numbers associated with each such author. This would eliminate the need for many of the "Author Inquiry" type questions and thereby reserve computer time for other types of searches.

c. Distribution to all users of search questions and answers of general interest.

d. Adoption of the Selective Dissemination of Information Plan under which pre-developed questions reflecting user interests would be put to the computer at least monthly in order to provide the individual user with a report of newly accessioned documents in the areas of his prime interests.

e. The storage of prime or recurring questions along with the accession numbers of the most relevant documents selected by the users from the total search results. Such refined search output could be inexpensively stored in a separate category.

f. A delete program should be written as quickly as possible in order to correct the few errors that have occurred in computer input and in order to make it possible to edit out and revise descriptors. Another computer program

Covington & Burling
16 November 1967

Page 5

should be prepared to automatically provide a listing of every new descriptor added to the computer since the date of publication of the Permuted Authority List and Thesaurus. This new programming feature would provide an alphabetical Authority List of new search descriptors which do not appear in these currently available documents.

This is by no means a complete set of review questions but may assist in beginning the process of evaluating the Legal Information Storage and Retrieval System after this first 11 month period of development and operation.

E.O.L.

SOS:vap
Attachments