

CANCERLINE: A New NLM/NCI Data Base[†]

SAMUEL A. TANCREDI,* RICHARD H. AMACHER, and JOHN H. SCHNEIDER

Office of International Affairs, National Cancer Institute, National Institutes of Health,
Bethesda, Maryland 20014

BRUNO M. VASTA

Specialized Information Services, National Library of Medicine, Bethesda, Maryland 20014

Received January 26, 1976

This paper describes collaborative efforts between the National Cancer Institute (NCI) and the National Library of Medicine (NLM) for the development of a new data base called CANCERLINE. The scope, content, and future enhancements of this on-line file are discussed, as well as the plans for its use to disseminate cancer-related information via NLM's biomedical communications network.

INTRODUCTION

On December 23, 1971, the President of the United States signed into law the National Cancer Act of 1971¹ authorizing the Director of the National Cancer Institute (NCI) within the area of information processing and dissemination to:

"...collect, analyze, and disseminate all data useful in the prevention, diagnosis and treatment of cancer, including the establishment of an international cancer research data bank to collect, catalog, store and disseminate, insofar as possible, the results of cancer research undertaken in any country for the use of any person involved in cancer research in any country. . ."

As a direct outgrowth of this law, the International Cancer Research Data Bank (ICRDB) Program was established. Its primary objective is to actively promote and facilitate the exchange of information between biomedical scientists and the dissemination of information through cancer centers and other appropriate organizations on a world-wide basis.

In May 1972, the National Cancer Institute convened a group of scientists representing various disciplines for an International Cancer Research Data Bank Planning Conference at Airlie House in Warrenton, Va.² A panel of 26 participants met for the purpose of bringing together in one place a group of knowledgeable scientists to discuss plans, problems, and schedules for the development of the ICRDB Program. The participants represented six countries, the World Health Organization, the International Union Against Cancer, the International Agency for Research on Cancer, the United Nations Educational, Scientific, and Cultural Organization, the American Cancer Society, and various other institutions and federal agencies of the United States. The Panel recommended that several steps be taken by the NCI to implement the information mandate contained in the Act, including the utilization and expansion of existing information systems capabilities and the planning for an ICRDB Program to bridge the information flow between the cancer researcher and the practitioner.

In order to implement the recommendations of the Airlie House Panel, the NCI supported a contract, beginning in March 1973, which produced a detailed user study,³ a survey of existing information resources,⁴ a system description,⁵ and a number of other studies completed in July 1974. A further detailed plan for the ICRDB Program was prepared in-house and the Program began moving into an operational phase.⁶

One of the more significant developments which enhanced the operational phase of the ICRDB Program took place

during the spring of 1973 when the National Library of Medicine (NLM) and the NCI began a series of meetings designed to support the ICRDB Program in meeting its objectives to disseminate cancer research results on a world-wide basis. In December 1974, these collaborative efforts led to the signing of an interagency agreement for a series of tasks to be performed by NLM, including the placement of cancer-related information (abstracts, citations, current cancer project descriptions) collected and processed by the ICRDB Program into a data base in the NLM computer system, and the dissemination of this information to various users via NLM's biomedical communication network.⁷

CREATION OF THE CANCER DATA BASE

In the fall of 1974, a new experimental data base called CCALINE (for *Cancer Chemotherapy Abstracts On-Line*) was constructed in the NLM's on-line system and made available on a trial basis to all MEDLINE (MEDLARS On-Line) and TOXLINE (Toxicology Information On-Line) users. This pilot data base (later to be called CANCERLINE) was tested by more than 300 MEDLINE/TOXLINE users and enabled the ICRDB Program staff to begin preliminary evaluations of the usefulness of the 16000 *Cancer Chemotherapy Abstracts* contained within the data base at that time.

As a result of these evaluations, the ICRDB Program is currently working with various in-house programs and externally via contracts and interagency agreements in order to collect, process, and disseminate the world's latest cancer research information. The sources of data for the Program include descriptions of on-going cancer research through an interagency agreement with the Smithsonian Science Information Exchange (SSIE); summaries of clinical cancer therapy protocols; abstracts from meetings, journals, books, monographs, and reports; and information from special ICRDB supported centers outside the United States. The information gathered from these various sources will be processed by a computer support contractor and formatted for input to NLM's computer system using the new ELHILL 3 software and made available as CANCERLINE.

GENERAL CHARACTERISTICS

CANCERLINE is a computer-based system for on-line retrieval of cancer information jointly created by the ICRDB Program and the NLM. It is a free-text search system which can search on any word or combination of words in titles and abstracts, as well as index terms assigned to the articles abstracted in the data base. At present, CANCERLINE is available to more than 400 locations linked to the NLM computer system in medical schools, medical research insti-

[†] Presented at the 169th National Meeting of the American Chemical Society, Philadelphia, Pa., April 9, 1975.

* Author to whom correspondence should be addressed.

CANCERLINE

NOW

- 21,000 CANCER THERAPY ABSTRACTS 1967-1975
- 23,000 CARCINOGENESIS ABSTRACTS 1963-1975

IN PROCESS

- 5,000 CANCER THERAPY ABSTRACTS 1975
- 4,000 CARCINOGENESIS ABSTRACTS 1975
- 6,000 IMMUNOLOGY, BIOCHEMISTRY, AND OTHER CANCER BIOLOGY ABSTRACTS 1975-1976

YEARLY GROWTH

- 15,000 ABSTRACTS FROM PUBLISHED SOURCES

Figure 1. CANCERLINE: content and status.

tutions, regional medical libraries, cancer centers, and hospitals throughout the United States, and in several countries outside the United States. Users can request CANCERLINE searches at any center currently subscribing to MEDLINE, TOXLINE, and other on-line search services provided by the NLM. The output of CANCERLINE searches is available to the general scientific community.

CONTENT AND STATUS

CANCERLINE currently contains about 45 000 citations and abstracts concerned primarily with cancer therapy and chemical, physical, and viral carcinogenesis (Figure 1). The following secondary sources are presently included: (1) *Carcinogenesis Abstracts* and (2) *Cancer Therapy Abstracts*. Several thousand additional abstracts in virology, immunology, biochemistry, and cancer biology were added to the data base in early 1976 through the use of screening and abstracting contractors. The CANCERLINE data base has a projected growth rate of 15 000 documents per year. Through an agreement with SSIE referenced above, 10 000 descriptions of current cancer research projects are also present in the CANCERLINE data base in a file called CANCERPROS for on-line searching.

Since most cancer agents are chemical in nature, another cooperative effort is underway with NLM whereby all chemical substances mentioned in the CANCERLINE bibliographic records will be identified by the Chemical Abstracts Service (CAS) Registry Number. This unique nine-digit Registry Number will serve to pull together all synonyms, including trade names, trivial names, company compound names, and numbers via access to the NLM's Chemical Dictionary File On-Line, known as CHEMLINE.⁸

SEARCHING CANCERLINE

CANCERLINE is accessed by a large variety of type-writer-like terminals connected to the NLM's central computer facility by direct telephone lines or via a nationwide network using telephone lines and small computers with access nodes located in many of the major cities. CANCERLINE is searched by entering desired terms as they may appear in titles or abstracts of the documents within this data base. Special CANCERLINE training has now been incorporated into MEDLINE and TOXLINE training classes which are periodically scheduled at NLM. Instructions for using CANCERLINE, including search commands and strategies, are almost identical with those used for TOXLINE.^{9,10} This means that there is no controlled vocabulary governing the inclusion or exclusion of terms in CANCERLINE. Terms may be entered singly, or in combination, by means of the Boolean operators AND, OR, and AND NOT. Searches may be limited to specific years of publication, secondary sources, authors, languages, or primary journal titles. Bibliographic citations may be printed on-line at the user's terminal or off-line and mailed to the user from NLM.

DATA ELEMENT	ABBREV.	DIRECTLY SEARCHABLE	STANDARD PRINT COMMANDS		
			PRINT	PRINT FULL	PRINT DETAILED
ABSTRACT	AB			YES	YES
AUTHOR	AU	YES	YES	YES	YES
AUTHOR ADDRESS	AA			YES	YES
JOURNAL TITLE	JT	YES			YES
LANGUAGE	LA	YES		YES	YES
SOURCE	SO		YES	YES	YES
SOURCE IDENTIFICATION	SI	YES	YES	YES	YES
TEXT WORDS	TW	YES			
TITLE	TI		YES	YES	YES

Figure 2. CANCERLINE unit record.

FIRST 5 TITLES ("PRINT TITLE" COMMAND)

TI-BILATERAL ADENALECTOMY FOR ADVANCED PROSTATIC CARCINOMA.
 TI-ADENALECTOMY AND HYPOPHYSECTOMY IN ADVANCED PROSTATIC CARCINOMA.
 TI-CEREBELLAR ATAXIA DURING 5-FLUOROURACIL.
 TI-THE EFFECTS OF ADENALECTOMY AND/OR HYPOPHYSECTOMY IN ADVANCED PROSTATIC CARCINOMA.
 TI-HYPOPHYSECTOMY AND ADENALECTOMY FOR DISSEMINATED PROSTATIC CARCINOMA.

Figure 3.

CANCERLINE UNIT RECORD

There are nine basic data elements within any CANCERLINE record, five of which are directly searchable as shown in Figure 2. As is the case for TOXLINE and MEDLINE, there are various standard "PRINT" options available to the user. One may select a print format varying from a brief identification of author(s), title, and bibliographic source to a complete listing of all data elements, including the abstract.

SAMPLE SEARCH

A typical request for a literature search dealing with carcinoma of the prostate involving removal of the adrenal glands (adrenalectomy) would be run on CANCERLINE in the following manner. The basic free-text terms describing the search statement should include all variations of the key terms, i.e., prostate, carcinoma, adrenalectomy, etc. Therefore, using the term truncation feature, one can specify:

ALL PROSTAT:

This will effectively retrieve all terms having this root, including prostate, prostates, prostatic, etc.

The same general technique is then used for the terms *carcinoma* and *adrenalectomy* in order to retrieve all possible word variations. The search strategy thus becomes:

Search statements	Term	No. of postings
SS/1	ALL PROSTAT:	372
SS/2	ALL CARCINOM:	3419
SS/3	ALL ADENALECTOM:	226

By ANDing the three search statements, one obtains the number of documents containing all search parameters specified:

1 AND 2 AND 3

This indicates that 12 documents meet the search specifications and the user can scan the titles of some or all of these documents as shown in Figure 3. Once the user is convinced that the documents retrieved are indeed relevant, he can request the complete document(s) as shown in Figure 4.

CANCERLINE CHARGES

At the present time, charges of CANCERLINE searches vary from center to center, depending on the terminal connect

FULL FIRST RECORD
("PRINT FULL" COMMAND)

SI— CATH/72/1661
AU—SCHOONEES, R.
AU—SCHALCH, D.S.
AU—REYNOSO, G.
AU—MURPHY, G.P.
AA—ROSWELL PARK MEM. INST., BUFFALO, N.Y.
TI— BILATERAL ADRENALECTOMY FOR ADVANCED
PROSTATIC CARCINOMA
SO— J. UROL.: 108 (1): 123-125; 1972
LA— ENG
AB—BILATERAL ADRENALECTOMY WAS PERFORMED ON 25 PTS.
WITH METASTATIC PROSTATIC CARCINOMA, WITHIN 3 MOS.
FOLLOWING ORCHIECTOMY AND ENDOCRINETHERAPY.
THOSE PTS. WHO HAD NOT RESPONDED TO ORCHIECTOMY
AND ENDOCRINE TREATMENT IMPROVED QUICKLY
FOLLOWING ADRENALECTOMY. SIX PTS. WHO HAD
RESPONDED TO TREATMENT BY ORCHIECTOMY AND
HYPOPHYSECTOMY, BUT WHO HAD RELAPSED, DID NOT
RESPOND TO SUBSEQUENT ADRENALECTOMY. (20 REFS.)

Figure 4.

time, telephone line charges, the amount of search output, and charges that may be made for time spent by the search analyst. Off-line print-outs which are mailed to users cost 10¢ per page; a maximum charge of \$15.00 per search has been established by NLM.

FLOW OF INFORMATION TO USERS

The CANCERLINE data base will also be made available to three Cancer Information Dissemination and Analysis Centers (CIDACs) being planned by the ICRDB Program. The CIDACs will cover the following three areas of cancer research: (1) cancer detection, diagnosis, therapy, rehabilitation, and other clinical aspects; (2) chemical, environmental, and radiation carcinogenesis (plus epidemiology); (3) cancer virology, immunology, biochemistry, and other basic cancer biology.

The major role of the CIDACs will be to provide the scientific subject expertise required to produce various services and products, to optimize the quality and validity of the scientific content of each product, and to ensure that the content and format of each product are of maximum utility to the scientists that receive the products. They will analyze, evaluate, and repackage cancer research information for dissemination to individual cancer researchers and clinicians. The services and products will include but not be limited to custom searches, technical bulletins, and the Selective Dissemination of Information (SDI). Schneider¹¹ presented an in-depth evaluation of an experimental SDI system with 104

principal investigators of research grants supported by NCI participating in the project.

CONCLUSIONS

A detailed plan for the ICRDB Program has been prepared and the Program has moved into an operational phase with the implementation of an on-line system called CANCERLINE developed jointly with the National Library of Medicine. The ICRDB Program plans to evaluate the usefulness of the system for cancer researchers, clinicians, and other system users over the next several months in order that the data base may be optimized for such users. Implementation of the three CIDACs will play an important role in this optimization. When fully operational, the ICRDB Program will satisfy the congressional directives as specified in the National Cancer Act of 1971 and meet the ICRDB Program objective of exchanging information between scientists worldwide and disseminating such information to physicians for the public good.

LITERATURE CITED

- (1) 92nd Congress of the United States Public Law 92-218, S.1828, Dec 23, 1971. The National Cancer Act of 1971 (8 pp).
- (2) International Cancer Research Data Bank Planning Conference, May 1972 (23 pp).
- (3) Informatics, Inc., "NCI and External Users Study", Final Report TR-74-1585-18, NCI Contract N01-CO-35403, April 1, 1974 (available from NTIS, Springfield, Va., as PB-238-178: \$4.75 paper copy; \$2.25 microfiche).
- (4) Informatics, Inc., "Directory of Cancer-Related Information Resources", TR-74-1585-17, NCI Contract N01-CO-35403, March 29, 1974 (available from NTIS as PB-238-024: \$7.25 paper copy; \$2.25 microfiche).
- (5) Informatics, Inc., "International Cancer Research Data Bank System Description", TR-74-1584-24, NCI Contract N01-CO-35403, June 28, 1974 (available from NTIS as PB-237-996: \$6.25 paper copy; \$2.25 microfiche).
- (6) "Plans for the International Cancer Research Data Bank Program of the National Cancer Institute", June 14, 1974 (unpublished, 22 pp).
- (7) D. B. McCarn and J. Leiter, "On-Line Services in Medicine and Beyond", *Science*, **181**, 318-324 (July 1973).
- (8) B. M. Vasta, D. F. Walker, and R. Schultheisz, "Conversion of TOXLINE Chemical Dictionary File into CHEMLINE", presented at the 168th National Meeting of the American Chemical Society, Atlantic City, N.J., Sept 11, 1974.
- (9) H. M. Kissman and D. J. Hummel, "TOXICON - An On-Line Toxicology Information Service", *Chem. Technol.*, **2**, 727-730 (1973).
- (10) B. M. Vasta, "The Use of TOXLINE and CHEMLINE for Retrieving Drug Information", presented at the Drug Information Association Meeting, Oct 28, 1974, Williamsburg, Va.
- (11) J. H. Schneider, "Selective Dissemination and Indexing of Scientific Information", *Science*, **173**, 300-308 (July 23, 1971).