

Computer Readable Data Bases and the Future of Information Services*

Dr G. Bhattacharya

Documentation Research and Training Centre, Bangalore

Deals with the varieties, genesis, and growth of Computer-Readable Data Bases (CRDB). Explains some technical aspects of CRDB's; and also the CRDB-services. Refers to the need for CRDB-resource sharing through networking. Enumerates the roles of the agencies offering CRDB-services. Takes note of the impact that CRDB has created; and indicates the future trend relating to information services.

1 Introduction

During the years of World War II, the human civilization witnessed with great fear and admiration the immeasurable extent to which man's physical powers had extended. But there had always been a doubt about the extension of man's powers of mind. Inventions, in those years, contributed extensively to the store of information. But the store remained bewildered. The efficient and effective accessibility to this store of information remained largely strewn with many insoluble problems for many years. This phenomenon was appreciably recognized by the industrially developed nations. The problems associated with this phenomenon were admirably realized by them.

As a result, during the post-war period, the human civilization has witnessed a more or less adequately balanced advancement of new developments in which "instruments" to ensure accessibility to information has figured prominently. Today perhaps, we can safely say that the necessary instruments are there. If they are properly developed, integrated, and utilized, they will provide users of information with access to and command over the inherited wealth of pertinent information. The task has become more of integrating different outstanding developments falling within the purview of different species of specialists. The responsibility of facing this complex situation largely lies on the profession of information scientists.

* Paper presented at the USICA Seminar on 'Entering the Information Age, Bangalore, June 3-4, 1981'.

To meet the requirements of the situation, it appears, that the profession has to enhance the capacity of its organs of perception to identify pertinent developments, evaluate its potential impact, and to make all attempts to utilize those developments as economically as possible to achieve its ultimate goal of "promoting the utilization of information to ensure national developments," which is, indeed, the directive of the normative principle "Information is for use." For this purpose, no doubt, the profession has to emerge from its traditional outlook and activities.

Science has provided us with the mechanism of speedy communication. One of its greatest contributions is the recorded information—that is, record of information in a communicable mode on a material surface fit for physical handling, transport across space, and preservation through time. It is only this recorded form that has made our inheritance of information meaningful. For, the manipulation, extraction, evolution, and endurance of existing information is guaranteed only by its recorded form. Man has developed his capability of extending the store of his recorded information enormously. But, he has not yet acquired the adequate capability of consulting the bulk of recorded information created by him. This is, indeed very important question; for, it involves the entire process by which man expects to profit by his inheritance of acquired information.

The recorded information has been largely responsible for the speedy development of the human civilization. But with the increase of

its bulk, the art and science of recording, and the material surface of recording have been gradually realized to be sources of great problems obstructing man's profit by his inheritance of acquired information. Search for solutions to these problems has led to a great evolution of the form of recorded information. Recording of information has reached a form and stage that bear great promises for the future. "Computer Readable Data Base" (CRDB) is a term that may be used to denote this evolved form and stage of recorded information. A humble attempt has been made here to trace the impact of CRDB on the future of information services. Today, for obvious reasons, developing countries can only take note of the experience of the developed countries, so far as the impact of a CRDB is concerned. There is, of course, a scope of imagination also. If we try to generalize the experience of the developed countries, there emerges a picture that was beyond imagination even two decades ago. The purpose of this paper is to furnish a brief outline of this picture. The primary focus of this picture is the bibliographic CRDB.

2 Computer Readable Data Base (CRDB)

A Computer Readable Data Base (CRDB) is an organized collection of information in computer readable form. CRDB's may be categorized on the basis of several characteristics. The following table furnishes a fairly comprehensive idea about those characteristics:

(1) On the basis of the types of data elements forming the records of the file.

Bibliographic—

e. g. MARC II data base of the Library of Congress (USA)
CA Condensates of the Chemical Abstract Service

Bibliographic-related—

- e. g. CASIA (Chemical Abstracts Subject Index Alerts) which contains subject index terms and posting that consists of Chemical Abstracts citation numbers).

Natural Language Text—

Full Text—

- e. g. *System 50 for State Statutes of Aspen Systems Corporation*

Summarized Text—

- e. g. *New York Times Information Bank of New York Times, Inc.*

Numerical—

- e. g. *U. S. Census Tapes*

Representational—

- e. g. *CAS Registry Structure*

(2) On the basis of the status of the producing agency.

Produced by Government Agencies.

- e. g. MEDLARS (Medical Literature Analysis and Retrieval System) tapes produced by the National Library of Medicine (USA)
- MARC II Tapes of the Library of Congress (USA)
- ERIC (Educational Resources Information Centre) Tapes of the National Institute of Education (USA)
- DDC (Defence Documentation Centre) Tapes of the Department of Defence (USA)'s Defence Documentation Centre
- GRA (Government Research Announcement) of the National Technical Information Service (NTIS) (USA)
- STAR (Scientific and Technical Aerospace Reports) Tapes, produced by the National Aeronautics and Space Administration (USA)

Produced by Private Agencies

Produced by Not-for-Profit Organizations

- e. g. SPIN (Searchable Physics Information Notices) Tapes of the American Institute of Physics.
- BA Previews (Biological Abstract Previews) of Biosciences Information Service.
- CA Condensates of Chemical Abstract Service.
- PATELL (Psychological Abstracts Tape Edition Lease License) of the American Psychological Association.
- COMPENDEX (Computerized Engineering Index) of Engineering Index, Inc.
- METADEX (Metals Abstract Index) of American Society for Metals.

Produced by Profit-making Organizations

- e. g. SCI (Science Citation Index) Tape ; and SSCI (Social Science Citation Index) produced by the Institute of Scientific Information.
- Excerpta Medica, produced by the Excerpta Medica Foundation
- F & S Index of Corporations and Industries, produced by Predicasts, Inc.
- New York Times Information Bank of the New York Times.

(3) On the basis of the subject matter covered by the CRDB's

Discipline Oriented—

- e. g. CA Condensates
 POST (Polymer Science and Technology)
 PATELL
 MEDLARS

Mission Oriented—

- e. g. INIS (International Nuclear Information System) Tape produced by the International Atomic Energy Agency.
 STAR

Problem Oriented—

- e. g. HEEP (Abstracts of Health Effects of Environmental pollutants).
 PIP (Pollution Information Project) produced by the National Science Library of Canada.

Inter-disciplinary—

- e. g. CBAC (Chemical and Biological Activities)

Multi-disciplinary—

- e. g. SCI
 SSCI
 MARC II

In addition to the characteristics mentioned in the table, there are others that affect the quality, timeliness and thoroughness of search results and the cost of processing. They are often used to characterize CRDB's. The following are some of those characteristics :

- (a) Kinds of primary sources of information included in the data base ;
- (b) Extent of coverage of those sources ;
- (c) Time-lag of coverage ;
- (d) Indexing policy employed ;
- (e) Extent of contents-description included ;
- (f) The Nature of search/display elements ;
- (g) Size and growth rate ;
- (h) Frequency of issue and update ; and
- (i) Existence or not of the corresponding hard copies.

3 Genesis and growth of CRDB's

Whenever an organized collection of information is converted into computer-readable form, a CRDB comes into existence. Between the late 1960's and early 1970's a considerable number of CRDB's came into existence. The majority of those were created as by-products of other activities. Quite a few were created because computers had proved to be economic effective tools for producing primary and secondary publications. Consequently, some were created because computer readable data were needed for computerized process control or production control of publications. Others were created for computer-aided typesetting. Very few CRDB's were created specifically for information retrieval. The reason was that the cost of data input was high, and it was difficult to justify for the purpose of information retrieval alone. It was simultaneously realized that

a CRDB was considerably flexible and could serve many desirable functions. They were readily amenable to reorganization, merger with other CRDB's, reformatting, and repackaging.

Today, the by-product aspect of CRDB's is no more of much significance. Many publishers have adopted a "CRDB-approach" to management of their processing system and distribution of their products.

The number of CRDB's existing today is far more than at any time in the past. It is now estimated that there are in excess of 500 CRDBs. Many more will be in existence in the near future. CRDB's are there in almost all of the major fields of science and technology. There are CRDB's covering news articles, legal cases and statutes, drug and poison information, etc. Efforts are under way to generate CRDBs that would provide community service type information—such as, consumer, day care, legal aid, recreational and leisure time activities information.

4. General technical aspects of CRDB's

In the case of a bibliographic CRDB, some of the items of information that are used to describe a source of information are: Names of authors and/or document associates, Title, Edition statement, Place of publication, Name of publishers, Date of publication, Collation statement, Name of series; Title of the host periodical, Locus statement, Subject indicative statement, Abstract/Annotation, etc. These items of information are the building blocks of CRDB's; and they are called DATA ELEMENTS. Pertinent data elements relating to a source of information is called a RECORD. An organized collection of records on a tape or other media together constitutes a FILE. A statement specifying the elements that are present, their location in the record, and the

length (number of alphanumeric characters) of the data content of the data element is referred to by the term 'DIRECTORY', often a directory is associated with each record. The arrangement of data contents, data elements, and of directory information for a record is referred to as the RECORD FORMAT. The arrangement of the records on a tape or other media is referred to as the FILE STRUCTURE or FILE FORMAT. It is possible to search specified data elements within an entire record. For this purpose, either the elements are identified or tagged with unique codes; or the position of an element within a record specifies the type of the element it is. When the records are arranged in a serial sequence and each record is given an identification number, the file is called a SEQUENTIAL FILE. When the searchable data elements are arranged with all their identification numbers, the file is called an INVERTED FILE. In an ON-LINE SYSTEM, the user is in direct communication with the central processing unit of the computer through a terminal. When there is a two-way communication between the user and the computer, and the time for response by the computer is immediate, the system is called an INTERACTIVE SYSTEM. In a BATCH PROCESSING SYSTEM, multiple jobs or search questions are taken together and run at one time. A RETROSPECTIVE SEARCH is one in which the search question is run against files containing records pertaining to sources of information that are not comparatively recent. The question is usually run once against a collection of many issues of a CRDB. A CURRENT AWARENESS SEARCH is one in which the search question(s) is/are run against a file containing records pertaining to sources of information that are comparatively recent. As a result, it is normally the same set of questions

(current awareness profile) that is run many times—each time against a different issue of the data base.

5. CRDB-Services

The information services that are most often generated from CRDB's are as follows :

- (1) SDI (Selective Dissemination of Information) service ; and
- (2) Retrospective Search Service.

Another information service known as "Private Library Service" is also based on CRDB's, and it is offered by a few organizations. A CRDB-search service naturally generates the demand for the relevant documents identified through the search. In this sense, now a days, "Document Delivery Service" is closely related to CRDB-search services. This service also is rendered by a few centres that process CRDB's. The following sections briefly deal with these CRDB-services.

51. SDI-Service

The SDI-Service is meant to keep the user up to date with the published documents in his field of interest. The user profile is usually designed to be as complete as possible to achieve high recall. The same profile is run over and over against new issues of the CRDB. Of course, whenever necessary, the profile is modified to suit the changes in user-interests. The SDI service is one of the most successful information services developed in the past decade. It has also been the primary use to which CRDB's were put in the early years of their development.

52. Retrospective Search Service

The purpose of Retrospective Search Service may be to provide the user with one or more

relevant references to help him to become acquainted with a topic or to find the answer to a specific question ; or with a fairly comprehensive coverage of references on a particular topic. The searches are conducted on demand in retrospective files. The use of on-line retrospective search services has grown considerably.

53. Private Library Service

A Private Library Service consists of offering to the user the facility of creating his own CRDB. This is done by storing in a user's own tape or disk file the output of his SDI runs. The file can be augmented by incorporating in it other records of the user's choice ; or it can be reduced by deleting records according to the user's instruction. Such a file is essentially tailored to the specific needs of the user.

54. Document Delivery Service

The Document Delivery Service is not yet widely offered by the centres that process CRDB's. One of the reasons for this is that very few CRDB processing centres handle document acquisition. Besides, hardly any closer ties exist between processing organizations and libraries. Since 1975, ISI and NTIS have been offering document delivery service for documents cited in their own CRDB's.

6. CRDB-Resource sharing through networking

The CRDB-services today have generated an economical necessity for CRDB-resource sharing. This is mostly due to the increased number of CRDB's, their voluminous size, and the associated cost of operation. It is no more possible to think of CRDB-services, whether for today or for the future, without thinking the need for and the advantages of CRDB-resource sharing through networking. CRDB-

resource sharing takes place mostly by remote accessing of CRDB's through communication network. No CRDB-processing centre can afford to process and provide services from all the available CRDB's. In spite of all that has been presented earlier, it is also well realized that

- (1) The cost of generating CRDB's and of producing them is quite expensive; and
- (2) The cost of establishing and maintaining CRDB-processing/searching activities is also high as it involves considerable investment in purchase/lease/licensing; royalty and access fees; materials and equipment; machine time, communication; and personnel expenses.

Besides, the cost of preparing, negotiating and conducting searches also is high.

CRDB-resource sharing through networking offers substantial advantages in this regard. For example, it can

- (1) Ensure availability of resources to a much larger community;
- (2) Reduce the cost of CRDB-searches as a result of distributing fixed costs over a larger base;
- (3) Effect reduction of number of skilled personnel needed for processing CRDB's;
- (4) Ensure accumulation of wider variety of experiences and "know-how" in CRDB-use;
- (5) Develop impetus towards standardization of CRDB-formats, element definition formats for search strategies, access procedures, etc;

- (6) Ensure availability of more resources at a single location; and
- (7) Ensure accessibility to CRDB's that individual user organizations would be unlikely to process internally because of low demand within the organization.

CRDB-resource sharing is largely done through communications networks. This is mostly confined to on-line systems. Besides, there exist other types of CRDB-resource sharing. For example, centres that process CRDB's themselves and provide services to internal and/or external clientele often require services for their own clientele from CRDB's that are processed in other centres. In such cases, two centres may exchange services or sell services to each other. Centres that provide their own SDI and retrospective search services in batch mode often function as middlemen in accessing on-line services for their clientele. They function in a referral capacity also in directing their clientele to the appropriate sources.

7 Agencies offering CRDB-services

The end-user of a CRDB-service, if he so wishes, can get it directly. Otherwise, he can get it through "Centres" or "Brokers". A centre, in this context, is an organization that acquires and processes CRDB's and provides services to users. The users may belong to the organization itself, or they may be outsiders. A broker may be a person or an organization that searches CRDB's on-line at another location; or purchases services from another centre, for its own customers. A broker does not possess CRDB's, but provides search services from them. Centres that operate computer-based information services as well as organizations that make use of on-line informa-

tion services, in general, provide an "intermediary" between a user and the computer-based system. The intermediary handles the intellectual tasks of selecting the appropriate system and CRDB's for the user's question; negotiating the search question with the user; developing the query or profile with an effective search strategy; conducting the search; and possibly evaluating the output.

8 Impact of CRDB's

What has been presented in the earlier sections is an outline of the experience of the developed countries in relation to CRDB's. The impact that CRDB's have created may perhaps be summarized as follows:

Today, far more users are receiving search services from CRDB's than at anytime in the past. A large number of CRDB's corresponding to many of the most heavily searched abstracting and indexing services now exist; and those also are heavily searched. This fact is quite significant. There are considerable data available now to assert that the level of use of CRDB's is rising rapidly.

It has been shown quite convincingly that the growth of CRDB's has

- (1) improved the availability of information sources;
- (2) overcome the barrier of geographical distance;
- (3) made information sources as readily accessible in a small community as they are in a major city; and
- (4) significantly reduced the cost of access to the sources of information.

One of the most successful information services developed in the past decade is the SDI service. The factors that have led to this development are:

- (1) increasing availability of computers;
- (2) the automatic generation of data bases through computer typesetting;
- (3) expansion of the universe of documents;
- (4) the increasing cost of labour intensive information services.

Upto the middle of 1970's only a few CRDB's had sufficient number of records to justify a worthwhile retrospective search. Within two years the situation changed. Today, the use of on-line services for retrospective searching has grown considerably.

The private library service based on CRDB's has become a very useful and effective entity of great attraction to individuals and institutions.

It is well recognized that a computer is not a substitute as yet for intellectual activities. In information retrieval, the intellectual aspects of searching remain the prerogative of the searcher. The intellectual role of an intermediary between a user and the CRDB's has been well recognised.

Very often a user completes a successful search only to be frustrated by the inability to readily obtain needed documents. This frustration has initiated concerted efforts for solving the document delivery problem. For example, CRDB-producers are thinking in terms of maintaining copies of all documents covered by them. Some are thinking of developing on-line searching facilities of one or more union lists of periodical publications with holdings information. Others are thinking of establishing national centres of periodical publications functioning as central depositories. These lines of thinking are expected to contribute to the development of a nation-wide resource sharing through networking; solution of legal

problems relating to document copying ; development of low cost facsimile transmission devices and their extensive use ; and reduction of communication rates for information transmission.

CRDB-services have made a direct impact on research and technical libraries in several ways. They have affected the acquisition policy of these types of libraries. They have pointed out the non-use of some periodicals and the need for others. They have affected the interlibrary loan traffic depending upon the correspondence between the library's collections and the retrieved citations from CRDB-searches. In many cases, the library has expanded or deepened its information services by offering services from the CRDB's it processes. In others, the library has been offering CRDB-services to its clientele by functioning as an intermediary preparing search questions and processing them via an on-line service or through another centre. Still in others, the library is functioning as a referral centre directing its customers to the appropriate CRDB's and service centres.

Accessibility through CRDB's has been described as "Electronic Accessibility". It is now claimed that the electronic accessibility of information sources is improving rapidly as the accessibility of printed sources is declining. Besides, the cost of electronic access is falling rapidly, as the cost of printed access is climbing. The most important point in this respect is that cost and accessibility through electronics will continue to improve, while cost and accessibility through print on paper can only get worse and worse.

The impact that CRDB's have created has drawn the encouraging attention of the national governments of many developed countries. They have specifically expressed their concern

in regard to making CRDB-services more economic, efficient and effective. For this purpose, they have recognized the following :

- (1) Need for training to work with CRDB's ;
- (2) Need for pertinent research in information and communication sciences ;
- (3) Need for expanding CRDB contents to serve new constituencies ;
- (4) Need for cooperation with private sectors in the generation and use of CRDB's ;
- (5) Need for development of resource locators, and improvement of document delivery system ;
- (6) Need for working towards a reduced rate for telecommunications for information transfer ; and
- (7) Need for CRDB-resource sharing via networks.

9 Conclusion

It is evident from the earlier presentation that CRDB's, their services and their associated technologies have set up a new trend towards many new possibilities of great consequences. Of course, there are many problems that would warrant effective solutions before we witness any such possibility in its perfect recognizable shape.

But there are strong indications that in the future there will be considerable improvements in all aspects associated with CRDB's leading to solutions of those problems. There will be more involvement of librarians, information scientists in CRDB-services. Such services will be made available in public libraries, as well as in academic, research, industrial, and administrative organizations. The techniques of computational linguistics, automatic contents

analysis, and pattern recognition will be employed on a larger scale. The aspect of user-computer interaction will receive more attention. Perhaps, natural language communication will make the use of the computer system easier.

Where all these developments will ultimately lead us to? It is, indeed, a very serious question. We have nothing more to be surprised if we see tomorrow that what we know as documents in their present form today, have all vanished. All documents of tomorrow have taken the form of magnetic records. They are being composed, published, disseminated, and used in a completely electronic mode. Paper is non-existent in this communications environment. The situation is that of a completely paperless system.

When we, the librarians, imagine of a completely paperless system, our thinking about the survival of libraries and about our roles and functions rests in the form of big questions only, without any satisfactory immediate answers.

REFERENCES

This paper consolidates works and views of many authors. In many cases, the same information and the same views have been expressed by many authors.

No attempt has been made here to cite the authors along with their works and views in the text. Several good attempts have been made to review the works in this area. The following is a list of some important documents in which the authors themselves have cited relevant documents quite extensively :

1. BRANSCOMB (Lewis M). Electronic library. *Journal of Communication*. 1981, Winter; 143-150.
2. CHERRY (Susan Spaeth). Telereference : The new TV Information Systems. *American Libraries*. 1983, February, 94-110.
3. DOWLIN (Kenneth E). Electronic library. *Library Journal*. 1980, November 1; 2265-2269.
4. LANCASTER (F Wilfrid). Whither libraries? or, Wither libraries. *College and Research Libraries* 39; 1978; 345-357.
5. TAYLOR (Robert S). Reminiscing about the future; Professional education and the information environment. *Library Journal*. 1979, September 15; 1871-1875.
6. WILLIAMS (Martha E). Impact of machine readable data bases on library and information services. *Information Processing and Management*. 13; 1977; 95-107.
7. WILLIAMS (Martha E.). Use of machine readable data bases. *Annual Review of Information Science and Technology*. 9; 1974; Chap. 7.