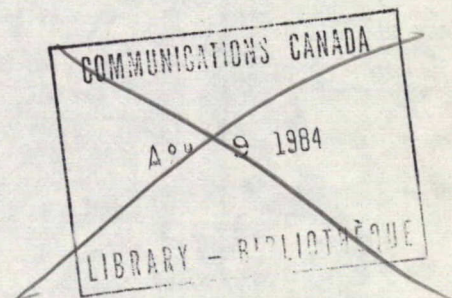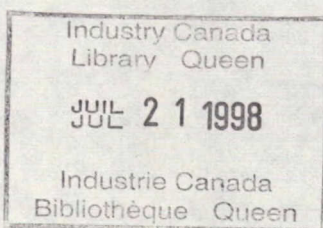FIRST DRAFT

RESTRICTED CIRCULATION

/ ② Report on

PUBLIC ON-LINE INFORMATION RETRIEVAL SERVICES /

Report to the

Working Group Economic Aspects and

Working Group Sovereignty Aspects

Interdepartmental Task Force on Transborder Data Flows

P.K. Neogi ①

J.V. Knoppers

Department of Communications

26 January, 1982

Contents

Public On-line Information Retrieval Services

Note 1:

1. Unforeseen delays in data collection and some logistics problems with the Appendices (Nos. 1, 2, 3 and 4 consist of rather voluminous information in tabular form) prevented the completion of the entire report for distribution at this time. The total length of this report will exceed 200 pages.

2. The major results and findings of the analytical work, which are detailed in the appendices, have however, been incorporated in the body of the draft report.

3. This draft report is being circulated simultaneously to the members of the TBDF Task Force Steering Committee, to the members of the Working Groups on Economic Aspects and Sovereignty Aspects and to industry. Upon receipt of written comments from the members of these two Working Groups chapter E (Conclusions) will be finalized. A final draft of the complete report, together with all the Appendices, will then be prepared and distributed for comment.

4. In the meantime, the attached Compendium of Major Issues and Findings serves as a working summary.

5. The Working Groups on Economic Aspects and Sovereignty Aspects have combined their work for this particular project to ensure that adequate

attention is given to both the economic and cultural components of online databases and retrieval services.

Compendium of Major Issues and Findings

1. Because of the lack of previous Canadian work in this area, at the start of the project it was necessary to:

   a. define a framework for analysis of the information services industry and determine which existing or readily available sources of information could be utilized to provide an overview on the nature, structure and scope of this industry;

   b. decide what additional data collection was feasible in the short term including interviews;

   c. indentify economic, sovereignty and cultural concerns of a TBDF nature that warrant investigation and analysis.

2. This report contains the first systematic analysis of the Public Online Information Retrieval Services "industry" in Canada. Its purpose is to study the industry in the Canadian and U.S. context, and to identify issues of concern.

3. Within the limitations of the existing information and data, this report:

   a. defines the Public Online Information Retrieval Services industry and its boundaries;

b. studies the structure of the industry and identifies three distinct sub-markets;

c. makes order-of-magnitude estimates regarding the size and growth trends of the market in the U.S. and Canada;

d. analyzes the relative sizes and growth trends of the sub-markets;

e. attempt to identify the nature and direction of transborder flows on online information services;

f. identifies economic cultural/sovereignty and legal/institutional issues of concern.

4. The public information services industry for the purpose of this report and the limits in time and resource of Phase I, has been defined <u>to include only the public use of online data bases in the information retrieval mode</u>. Intra-firm and closed user group application, the bulk of which are in the transaction processing mode, have been excluded for now. Public on-line information services are seen to be only a part of the much larger network information services activity.

5. The industry consists of five principal groups of players. They are

a. the information providers, data base producers and data base distributors or vendors, who constitute the supply side of the industry;

b. the information brokers and information buyers or end users, who constitute the demand side.

Telecommunications carriers play a key support role by providing the data networks linking the user terminal to the central computers of the vendors. They have been considered as a supplier of services, rather than an integal component of the industry.

6. The worldwide market in 1981 was over $1 billion (U.S.) and it is growing rapidly. The recently published Cuadra Directory of Online Databases listed 965 databases, 512 information producers and 170 vendors. Compared to the numbers listed in the same directory a year ago, there was a 57% increase in the number of databases and a 83% increase in the number of vendors.

7. The market is predominantly a business one and likely to remain so in the medium term, although mass consumers may eventually become significant users of on-line databases.

8. The market can be subdivided, by online database type, into two distinct sub-markets:

   a. Reference Databases, used mainly by libraries and information specialists, which constitute about 40% of the total number of databases, but account for only some 10% of the revenues to vendors;

   b. Source Databases, which include Economic, Financial, Credit, Marketing and Demographic databases, are the ones most heavily used by business firms. This is by far the largest sub-market in terms of revenues (some 75% of the total), and one of the fastest growing.

9. The report highlights the lack of hard data (usage volumes and revenues, absolute and trends) regarding the demand side of this industry generally, and for the Canadian marketplace in particular. Any attempts at a further or more precise analysis of market shares, sub-market growth trends and transborder flows of information services are currently limited by data availability.

10. The U.S. domestic market constitutes the dominant share of the worldwide market for public online information services. For 1980, market revenues to database vendors are estimated to have been in the range U.S. $600-900m. The nominal compound annual growth rate is estimated to be in the range 30-38%.

11. The Canadian market for 1980 is assumed to be 5-7% of the U.S. one, with a range of Canadian $20-45m. The nominal growth rate is assumed to be 30%. This would produce a 1985 market size range of Canadian $75-167m.

12. The reference database sub-market in Canada has a very high proportion of imports, both with respect to databases and services. The sub-market size for 1980 is estimated to be $5-7m, with a possible negative trade balance of $4-5m.

13. Credit databases are a major area of potential concern, because of privacy and sovereignty, as well as economic considerations.

14. Given the existence of CANSIM and the activities of I.P. Sharp Associates in particular the balance of trade for socio-economic databases could even be positive.

15. The balance of trade for full text legal databases may be even, but is probably negative for news databases.

16. The overall balance of trade between Canada and the U.S. in packaged databases and public information services probably lies in favour of the U.S. However, even if imports accounted for 60% of the domestic Canadian market and there were no exports (an extreme scenario), the negative trade balance would still be only $12-$27m. This is only 3-5% of the $530m of EDP services estimated (by one source) to have been imported by Canadian users in 1980.

17. Cultural and sovereignty concerns are increasing as the industry continues its rapid growth.

18. The lack of a clear legal framework for the information industry is considered by industry to form a major impediment to its development.

19. A major reorganization (editing) of the report is in progress, suggestion from readers are welcomed.

<u>Public On-Line Information Retrieval Services</u>

<u>Note 2:</u>

<u>Canadian Computer Service Industry and Service Bureau Revenues</u>

There is an apparent discrepancy between the Commercial Service Bureau Industry paper and the Public Online Information Retrieval Services paper in the estimates used for total revenues of Computer Service bureaus. The first paper (p. 3) derives an estimate of $500m in 1979. The second paper (p. 54) uses a 1980 estimate which ranges from a low value of $463m (for the top 35 Canadian Computer service bureaus), to a high value of $827 (CADAPSO estimate based on Evans Research Corporation figures). The relevant statistics and derived estimates for 1978-80 are given in Table 1, and explained below.

It should be noted that service bureaus are a subset of the total number of firms comprising the computer service industry. Because of coverage problems related to methodology, the Statcan Computer Service Industry report fails to cover all firms engaging in commercial data processing activities. Therefore the total operating revenues for the computer services industry given in by Stats Can. 63-222 have to be adjusted upwards; the DOC adjusted estimates (2) are therefore 26.3% higher in 1978, 22.1% in 1979 and 19.6% in 1980 while the Evans estimates are 10-15% higher than the DOC ones.

The $500m service bureaus revenue estimate for 1979 in the Commercial Service Bureau Industry paper was derived as follows:

The data processing services revenues of $357m were increased by $36.8m to adjust for hardware equipment firms (IBM, etc.) to obtain a total of $374m (this is a misprint, the correct value should be $394m). This revenue was allocated to service bureaus, which were assumed to obtain 75% of their total revenues from data processing services. The total services bureau revenues are, therefore $374/0.75 = $499m, or, more correctly, $394/0.75 = $525m in 1979.

The corresponding figure for 1980 is $467/0.75 = $622m.

The difference between the $463m estimated by Evans for the top 35 service bureaus, and the $622m calculated above, is due to a number of factors, principally differing coverage and the mix of data processing with other EDP activities for service bureaus. The estimate of $622m does, however, lie well within the range $463m - $827m; in fact, the midpoint of the range is $645m. The service bureau revenue estimate is not, however, a consistent measure of the data processing services activity. It is perhaps this activity which should be compared with the estimate of the public online information services market.

The apparent discrepancy serves to underscore the need for an agreed upon methodology for analysis and the fact that the various sources of data are based on differing definitions of the industry.

TABLE 1.   CANADIAN COMPUTER SERVICE INDUSTRY REVENUES

Comparison of Statcan, LAS and Evans Estimates
(Revenues in $m)

Item	1978 STC	1978 Evans	1979 STC	1979 Evans	1980 STC	1980 Evans	COMMENTS
A. Computer Service Industry							
1. STC unadjusted							
– Total Operating	532		638		820		– Should use Operating
– Total	539		642		824		rather than total rev.
2. Total Computing Services (LAS)	672 (26.3%)		779 (22.1%)		981 (19.6%)		– Derived from STC by adjustments for lack of coverage; % difference shown
3. Evans Estimate		776 (15.5%)		834 (14.8%)		1060 (8.1%)	– % difference with LAS estimates shown
B. Service Bureau Activity							
1. Evans (Top 35 Firms, less than total)		329 (42.4%)		389 (43.5%)		463 (43.7%)	– Expressed as % of total CSI revenues
2. STC-Inadjusted							
o Processing Services	314 (59.0%)		357 (55.9%)		433 (52.8%)		– Processing and Software/Systems revenues expressed as % of total Operating Revenues
o Software/Systems Services	112 (21-1%)		136 (21.3%)		208 (25.4%)		
3. Processing Services – RH adjusted	348		374* (394)		467?		– Adjusted by adding estimate for IBM SB
4. Service Bureaus	464		499*		622?		– Assumed that Proc. Serv – 0.75XTOR
5. LAS ADJUSTED							
– Computer Processing	432 (64.3%)		487 (62.5%)		571 (58.2%)		
– Software/Systems Services	161 (24.0)		194 (24.9%)		293 (29.9%)		

Sources: 1.   Statistics Canada, Catalogue 63-222, Computer Service Industry 1978, 1979, 1980 (Prelim).

2.   EDP In-epth Reports, Evans Research Corporation, May 1980, March, April 1981.

3.   L.A. Shackleton Adjustments to STC Source Data, for Growth Model Estimates.

A. INTRODUCTION

1.1 Background

While some have been predicting the advent of the "information revolution" or the "electronic revolution" for a quarter of a century, it is only during the last few years that this "revolution" has become an imminent reality. Order of magnitude decreases in computer processing and storage costs combined with equally rapid technological advances in telecommunications and communication networks, have been the primary factors in the appearance of the "information industry".

Information has always been a saleable commodity and it has been used to improve the decision making process. The role of information has grown steadily in post-industrial societies, as an increasing share of the work force becomes information workers and an ever-increasing number of firms and individuals become consumers of information. For others, information is now considered to be the fourth factor of production or the fourth element to be managed. While discussion of these concepts is already in full flower and in some cases are actually being applied, the fact remains that information itself is growing at an exponential rate. Some estimates indicate that between 1970 and 1985, the world's stock will have grown in volume between four and seven times.

The growth is the result of a complex mix of demand and supply factors:

° the expansion of science (i.e. new fields of scientific inquiry);

- ° the linking of science to the new technologies and the increasing importance of technology to science;

- ° the increase in population (i.e. the effect of the post-war baby boom and lower infant mortality);

- ° greater literacy and more schooling (i.e. the expansionary boom in post-secondary education);

- ° the growing demand for news, entertainment, information and knowledge;

- ° the establishment of "real-time" communication and world-wide communication links of computer networks with greatly increased data flow capacity (i.e. telephone, cable, satellites, some of the latter having transmission rates of over 6 million bits per second per transponder);

- ° the quantum decreases in data processing and data storage costs.

In a wider context, the recent rapid evolutionary or revolutionary change of converging computer-communication technology, reductions in per unit costs and an increasing supply of data have by their very nature led to a tremendous increase in the volume and extensiveness of electronic data flows.

These data flows form the arteries of the "Global Village". As advances in computer-communications networks link nations more closely together, the question arises as to what effect this transformation has or may have on the nation-state. International organizations such as the Council of Europe and the OECD and a number of European nations have already formulated laws, regulations, guidelines, etc. dealing with particular aspects of transborder data flows (TBDF).

Much of this information is no longer handled by means of the traditional printed work. To remain useable and to be delivered where and when required, it is created, stored, manipulated, accessed and transmitted by electronic means. This major change in the way information can be collected, packaged and distributed has led to the emergence of a new industry based on the use of on-line data bases, huge banks of data or information that are processed, stored and retrieved electronically.

On February 26, 1981 the Minister of Communications established an Inter-Departmental Task Force to study the implications of rapidly increasing Transborder Data Flows (TBDF) for the Canadian economy and for Canada's sovereignty. The Task Force Steering Committee established three Working Groups, with respective responsibility for the analysis of the economic impacts of TBDF, the sovereignty impacts and the international environment.

The present project on Public On-line Information Retrieval Services is a joint undertaking of the Economic and Sovereignty Working Groups. It represents a combination of the work started under the Information Retrieval Services project of the Economic Working Group with that of the Cultural Aspects team of the Sovereignty Working Group. A single set of reports, of which the present one is the main working document, will be presented to both groups. The study team considers that this joint approach will yield a better understanding of the facts, issues and problems related to Public On-line Information Retrieval Services than two individual studies. This is because one must study the content of such services as well as usage, access and delivery since it is the content or the information that is being sold and bought in this industry,

not the technology as such. Content poses issues related to both economics and cultural sovereignty.

1.2 Purpose of the Report

The purpose of this report is to study the Public On-line Information Retrieval Services "industry" in the U.S. and Canadian context, and to identify issues of concern. The bulk of the analysis relates to the structure and economics of the "industry". Cultural/sovereignty and institutional/legal issues are also identified.

The tasks related to the economic analysis have been formulated as follows:

1) Define the industry and its boundaries:

2) Undertake a preliminary survey of the available literature and identify the data, if any, upon which quantitative analyses may be based;

3) Study the structure of the industry and identify distinct sub-markets if they exist;

4) Calculate order-of-magnitude estimates regarding the size and growth trends of the Public On-line Information Retrieval Services market in the U.S. and Canada;

5) Analyze the relative sizes and growth trends of the sub-markets, within the limitations of the existing data;

6) Identify the nature and direction of transborder flows of Information Services, to the extent possible from available information.

An attempt has also been made to identify the main Canadian database producers and vendors and compile an inventory of their service offerings. The producers and vendors identified are discussed in this report. The inventory of databases currently being accessed by Canadians, based upon existing directories and sources, is detailed in a companion document which forms an appendix to this main report.

In addition to analyzing the above types of "factual" information, the report examines concerns expressed by Canadian database producers, vendors and some end users. The purpose of this exercise is to identify the major issues and problems related to both the supply and the demand sides of the Public On-line Information Retrieval Services industry, and the impacts that TBDF is likely to have on both Canadian suppliers and users of such services.

1.3 Limitations of the Report

This report suffers from a number of limitations. These arise from the time and resource constraints under which it was prepared; the lack of meaningful demand data and the inability to collect such data by mounting a survey of users; and, finally, by the fact that it is the first extensive analysis of the Public On-line Information Retrieval Services industry in Canada. With the exception of the recently published Evans report on Public Databases (Ref. B12), there is no available document against which even the

broad conclusions and findings of this report regarding the Canadian industry can be compared and tested. Considerable further work will be required in this area before a comprehensive picture of the Canadian Public On-line Information Retrieval Services industry, consisting of both the demand and supply sides, can be created and the problems studied in-depth.

It should also be noted that Public On-line Information Retrieval Services is only one of the many subsets of the much larger and more extensive on-line network/transaction information services industry. However, a discussion of "information industry" as a whole cannot be undertaken without introducing the issue of privacy. As the federal government is currently in the process of formulating new principles and privacy rights for personal information under its control, it was considered premature for the Task Force to consider privacy issues at this stage.

Nevertheless, the Public On-line Information Services Industry even though it presents only a relatively small subset of the information industry presents a complex enough problem of analysis as is demonstrated by this report.

Finally, this report focuses on the public on-line service industry and in doing so does include estimates for total revenues of computer service bureaus. However, those wishing to study the latter should consult the Report on Data Processing.

B. INDUSTRY STRUCTURE AND TRENDS

2. Definition of Industry and Boundaries

The term "Information Services" must be given a workable definition and clear boundaries, before any meaningful measurement of this "industry" is possible. In particular, one must realize that such services constitute only a subset of the existing and potential uses of online databases.

The online mode of use is characterized by the following components:

a) user terminals, often intelligent, which provide input/output, communications and, often, some local processing capabilities;

b) a telecommunications network linking terminals to one or more central computers;

c) powerful timesharing computers with appropriate applications software capable of organizing large databases, flexible searching and retrieving the information sought by the user;

d) large databases stored online and accessed/queried interactively. (Data files used in a batch processing mode are therefore excluded).

The explosion in the use of online database during the last 10 years has been due to major improvements in the price/performance of computers, online storage technology, the emergence of reliable telecommunications networks designed for data transmission and the development of reliable database management applications software. The huge increase in the amount of source

information created and/or captured in electronic machine-readable form has also spurred the development of online databases. The increasing use of word processors in offices, the extensive use of computers to maintain organizational records and the automation of the print publishing process (using photocomposition and automated type-setting) means that an ever-increasing proportion of source information is now created in electronic form, rather than hard copy. Combined with the use of optical character recognition (OCR) for converting printed information into electronic form, this means that the cost of creating machine readable input for online data bases is greatly reduced, thereby removing an important economic barrier.

2.1 Information Retrieval and Transaction Processing

The use of online databases can be categorized into two broad modes: information retrieval and transaction processing. In the information retrieval mode, the end user is restricted to searching and retrieving information from the database. He cannot initiate a transaction which can change the content of the database; updating is rigidly controlled by the database vendor. All public online information retrieval services, which must provide widespread access, operate in this mode.

In the transaction processing mode the end user can initiate a transaction which, when processed, changes the status of the database. Systems used for inventory management, airline reservations, online banking, point-of-sale services and many other applications fall into this category. All transaction processing systems have secondary information retrieval capabilities, through the use of query transactions. Such systems have generally been used

for intra-company or closed user group applications; but with the spread of teleshopping, telebanking and EFTS (Electronic Funds Transfer Systems), the semi-public use of such systems is likely to become widespread over the next decade. The transaction processing mode probably forms the bulk of intra-company and closed user group applications, but this form of use does not normally result in market transactions. It is buried or dispersed in the total EDP, communications, library and other information related (e.g. marketing) expenditures of organizations. Expenditures associated with the transaction processing mode of use, within firms and closed user groups, could well be 5-10 times larger than those associated with public online information services in Canada.

This study will focus upon the information retrieval rather than the transaction processing mode of using online databases. The analysis of public online information services will be primarily concerned with the public use of online databases in this mode. This is because public transactional databases are still in their infancy, and there is very little available information in this area. It should be noted, however, that videotex based services, e.g. Telidon are being designed to provide transactional capabilities. In fact, the success of such services may very well depend upon their transactional and gateway capabilities. This point is discussed further in section 5 and Appendix 5.

2.2 Industry Structure

The public online information retrieval services "industry" consists of five principal groups of players: the information providers, database producers,

database distributors or vendors, information brokers and the users.
Telecommunications carriers play a key support role by providing the data
networks linking the user terminals to the central computers of the vendors, but
they will be considered a supplier of facilities and services to the "industry"
rather than an integral component.

a)   Information providers are the organizations which originate or create
     information. They include the learned societies, professional
     organizations, abstracting services, news services, research
     organizations and government agencies. The information may or may not
     be created in electronic machine-readable form. If the information
     provider creates the information in machine readable form, he may also
     become the database producer, either in a primary role or as a spin-off
     to a primary hard-copy product.

b)   Database producers are the organizations responsible for the creation
     and updating of databases in machine readable form, using information
     supplied by the providers, which could be themselves. A wide variety
     of organizations produce databases, including government agencies,
     research firms, universities, print publishers, financial institutions
     and database specialists. The Canadian government is the largest
     single producer of public access databases in Canada, but this is not
     so in the U.S. It should be noted, however, that many of the U.S.
     database producers and vendors, in their formative stages, relied
     heavily on contracts and funding provided by government agencies.

c) <u>Database distributors or vendors</u> are responsible for storing databases online, providing access/retrieval software and computer access via remote user terminals. The vendor buys from the producer the right to make the latter's database publicly available via the vendor's computer. The database producer may be paid in a variety of ways, such as a flat annual fee, royalties based on volume of usage, or both. In some special cases the database producer may pay the vendor for the privilege of making his database widely accessible in electronic form.

Some vendors are computer service bureaus providing timesharing services, who offer databases as a part of their total range of services to clients. Other vendors, like the search services, have as their main raison d'être the distribution of public databases. Some vendors, like I.P. Sharp Associates and Data Resources Inc. (DRI) operate their own private data networks. Almost all vendors now provide access via public data networks like Datapac, Telenet and Tymnet.

d) <u>Information brokers</u> act as an interface between the end user and the source of information, charging a fee for the service. Reference librarians have traditionally played this role, between end users and printed information in the past, and now the vendors of bibliographic databases, but they have not charged an explicit fee for the use of their services. A large number of fee-based information services have sprung up in the U.S. over the last 10-15 years, to cater to the requirements of the customer who needs specific information, but is unable or unwilling to acquire it directly from the database vendors.

There is an obvious analogy, and even overlap between such activities and traditional consultancy services.

e)   End users are the buyers of information services, the entities to whom the distributors market their databases, either directly or through information brokers. The end users come from government agencies, almost every segment of industry, financial institutions and professional service groups. Although the market is predominantly a business one and likely to remain so for the medium term, consumers also may eventually become significant users of online databases.

The information providers, database producers and vendors constitute the supply side of the Public Online Information Retrieval Services industry. Vertical relationships, formal or informal, are possible between the information providers and database producers, and also between the producers and vendors. An electronic publisher may undertake all three functions; examples would include the Globe and Mail, the New York Times and McGraw-Hill.

The end users and information brokers constitute the demand side of the industry. Since the end user of public online information retrieval services could be, potentially, any firm or even individual consumer, the demand side is difficult to survey or study. The problems faced in studying the demand side of this industry are somewhat analogous to those of studying computer use, except that there is even less hard Canadian data available in this case.

3. Basis for Quantitative Analysis

The first task related to any attempt at quantitative analysis must be to determine what data and statistics are available in this area, and their relevance to the proposed study. The following possible sources, which collect and disseminate a variety of EDP statistics on an annual or ongoing basis, were examined:

1. Statistics Canada

   The Computer Services Industry Report (Cat. 63-222, 1972-79);

2. Treasury Board

   Review of EDP and Telecommunications in the Government of Canada (Annual Reports);

3. Canadian Information Processing Society (CIPS)

   Canadian Computer Census (Annual publications, 1965-1980);

4. Evans Research Corporation of Canada

   EDP In-Depth Reports (Jan. 1979 - Oct. 1981 issues);

5. American Association of Data Processing Service Organizations (ADAPSO)

   1978 Annual Report;

6. Canadian Association of Data Processing Service Organizations (CADAPSO)

   Annual Surveys of Member Firms

The October 1981 (Vol. 11 No.2) issue of Evans' EDP In-Depth Reports is devoted to Public Databases. With the exception of this issue, however, examination of the data provided by the above publications showed that there was virtually no separate statistical coverage of information retrieval activities, although some attention has been given to the problems of developing, maintaining and using online management information systems (MIS). It was

therefore concluded that, with the exception cited above, none of the above sources, either individually or in aggregate, could provide the data necessary for an adequate quantitative evaluation of public information services in the Canadian or North American context.

However, several organizations now maintain and publish directories covering the products and services available in this area. The following four directories were consulted and the information provided analyzed.

1. Encyclopedia of Information Systems and Services (Ref. B1)

This Encyclopedia, edited by Kruzas and Schmittroth, is now in its 4th edition (1981). It provides comprehensive coverage of information products and their producers. More than 2030 organizations are described in this edition. The main inventory is arranged by parent organization name in alphabetical sequence. There are 22 indices which provide a detailed analysis of the contents of the Encyclopedia and the fields covered, 14 indices classify listed organizations by general type of activity or function (Database Producers, Data Collection and Analysis, Online Vendors/Telecommunications Networks, Abstracting and Indexing Services, etc.), and 8 indices provide access to entries through such specifics as database name, subject interests and geographical location.

Within the limitations of availability and appropriateness, the following items of information are given for each organization:

1. Name, Address and Telephone Number
2. Founding Date

3. Head of Unit

4. Staff

5. Related Organizations

6. Description of System or Service

7. Scope and/or Subject Matter

8. Input Sources

9. Holdings and Storage Media

10. Publications

11. Microform Products and Services

12. Computer-based Products and Services

13. Other Services

14. Clientele/Availability

15. Projected Publications and Services

16. Remarks and Addenda

17. Contact

2. Computer-Readable Databases: A Directory and Data Source Book (Ref. B2)

This directory, edited by Martha Williams et al., was published by the American Society for Information Science in 1979, updating the earlier publication of 1976. It was produced from data maintained in the "Database of Databases" by the Information Retrieval Research Laboratory (IRRL) at the University of Illinois at Urbana. All the 528 databases included in this directory fulfil three criteria:

i)   they are in computer-readable form;

ii)   they are publicly available; and

iii)   they are used for information retrieval purposes, or are available

through the major online vendors of information retrieval services.

The main inventory consists of a detailed description of each of the

528 databases, arranged in alphabetical sequence according to IRRL record name.

It is supplemented by 4 indices, which are ordered by database producer name,

processor (i.e. vendor) name, database name and subject matter.

For each specific database, the information available is grouped under

eight major headings, as follows:

1.   Basic Information

2.   Producer/Distributor/Generator Information

3.   Availability and Charges for Acquisition of Database Tapes

4.   Subject Matter and Scope of Data in Database

5.   Subject Analysis/Indexing Data

6.   Data Elements Present

7.   Database Services Offerred

8.   User Aids Available

3.   Directory of Online Databases (Ref. B3)

This directory, published quarterly by Cuadra Associates Inc., is

compiled and edited by Cuadra, Abels and Wanger. The present edition, Volume 3

Number 1, was published in the fall of 1981. The criteria applied in selecting

a database for inclusion in the directory are as follows:

i)   it must be available <u>online</u>;

ii)   it must be available to the public, or to organizations that can establish their eligibility through subscriptions or membership;

iii)   it must be accessible through an online service organization that is connected to one or more international telecommunications networks, and/or to networks that serve one country, or a limited set of countries.

The main inventory, arranged in alphabetical sequence by database name, <u>consists of 760 entries covering approximately 965 databases and distinctly named files within database families</u>; this number represents a <u>57% increase</u> in over the number of entries in Volume 2, Number 1, published in the fall of 1980. Four indices are provided, ordered by subject matter, producer name, online service name and database name. Addresses and telephone numbers are provided for the <u>512 producers and 170 online services</u> listed. The number of online services listed in the directory has increased in 12 months by 83%, from 93 in Volume 2, Number 1 through 135 in Volume 2, Number 3, to 170 in Volume 3, Number 1. The growth in the number of entries and online services primarily reflects an increase in new products/ services, but expanded scope and coverage of the directory has also added to the number of online services. The Cuadra Directory has been used as the basic source of information for the tables in Appendices 1 and 4.

The information for each database is provided under the following set of headings:

1. Name of Database

2. Type (2 level classification)

3. Subject

4. Producer

5. Online Service

6. Conditions for access

7. Content

8. Coverage

9. Updating

4. COIN: A Directory of Computerized Information in Canada (Ref. B4)

This directory, now in its second edition, provides extensive coverage
of public online databases currently being accessed by Canadians. Edited by
Sandra West, COIN is published by the Alberta Information Retrieval Association.
It is also available for online searching on the SPIRES system at the University
of Alberta in Edmonton. A database must satisfy three criteria to be included
in the directory. It must be:

    i)   publicly available in Canada;

    ii)   in machine readable form;

    iii)   contain statistical or bibliographic information.

Because of the restriction regarding public availability in Canada, the coverage
of COIN is less than that provided by Williams or Cuadra.

The main inventory consists of a detailed description of about 350
databases, arranged in alphabetical sequence by database name. Unlike Williams
and Cuadra a separate entry is given for each host system at which the database
is available. Thus ABI/INFORM, which is available through BRS, Infomart (SDC)

and DIALOG, generates 3 entries. The main directory is supplemented by a Keyword Index providing subject reference to the databases, and a Contact Index providing more detailed information about each search centre.

The following information is provided for each database entry:

1. File Name — Including database name, acronym and host system (vendor)

2. Type — Statistical or Bibliographic

3. Source — Database Producer

4. Contact — List of Contact Organizations in Canada

5. Abstract — Explains the subject coverage

6. Data Types — Description of data elements

7. File Size — Total number of records, as of a given date

8. Updating — Update frequency and number of records added per update

9. Time Span — Time period covered

10. Location — Address of host system

11. Storage — Computer storage medium, usually disc

12. Machine — Model and size of the host computer

13. Restrictions — Availability of the database to the public

14. Cost — Database charges as designated by the host system

15. Searchable — Online or Offline

Note that items 10 to 15 can vary from one host system to another. A detailed analysis of the information in the COIN Directory is provided in a companion report.

The above directories give a considerable amount of information regarding the supply side of the public information services industry. What is lacking, however, are statistics related to usage volumes and revenues; therefore, the directories cannot be used to construct a picture of the demand side. Industry studies, if available, must be used to fill this gap in the information.

The next step was to look for industry studies of the Online Database Services markets. Four such studies, by International Resources Development (IRD) Inc., Creative Strategies International (CSI), Link Resources, and Input Ltd., all published in 1980-81, were identified (See References C1, C2, C4 and C5). The first three studies deal with the U.S. market and the fourth with opportunities in the European market. With the exception of the Evans report cited earlier, no similar studies have been identified as yet which attempt to deal with the Canadian market. The IRD study was acquired. The Link study cannot be acquired on an individual basis and a subscription to the Online Database Planning Service ($15,000 annually) was not considered as cost justifiable for the purpose of this phase of the project. However, access to the main results of the study was obtained through the National Library and the Telidon project. Summary results from the LINK, CSI and Input studies have been published in the EDP Weekly.

Faced with this situation, it was decided that the most appropriate course of action would be as follows. A review of the size and growth trends of the U.S. market would be made using the industry studies available and any other relevant published material (e.g. Ref. A3, A7, A8). A "scale and lag" process, supplemented by whatever quantitative information could be collected from Canadian information producers and database vendors, would then be used to

derive order-of-magnitude estimates of the current Canadian market for public

information services. In attempting to project the Canadian market to 1985 and

1990, U.S. growth trends would be modified, where judged necessary, to take into

account the impact of purely Canadian factors.

4. <u>Classification of Online Databases and their Characteristics</u>

There are a number of ways in which one can classify databases. Databases vary by subject, scope, type of information, geographic and chronological coverage, frequency of updating and also by record/file structures, types of systems with which they can be used, intended uses and intended users. The classification scheme used in this report was proposed by Wanger and Landau (Ref. A5). It is also used for the Cuadra Directory of Online Databases (Ref. B3).

The Cuadra Associates classification scheme is a two-level one, based upon the type of information contained in a database. The primary distinction is made between (1) Reference databases (i.e. those that refer users to a primary source for complete information), and (2) Source databases (i.e. those that themselves contain the source or primary information). Within each of the two primary types, there is a secondary classification, as shown below.

4.1 <u>Classification Scheme</u>

1. <u>Reference Databases</u>

These databases refer or "point" users to another source (e.g. a document or an organization) for additional information or the complete text of a citation or abstract. All Reference databases are textual in nature. They are further sub-divided into:

1.1 Bibliographic

These databases contain citations and, sometimes abstracts of the printed literature, e.g. journal articles, reports, books, newspaper items, conference proceedings, dissertations. Most of the databases offerred by the search services (Dialog, SDC, BRS, CAN/OLE) fall into this category.

1.2 Referral

These contain references and, sometimes, abstracts or summaries of non-published information. They generally refer users to organizations, individuals, audio-visual materials and other non-print media for further information.

2. Source Databases

These databases contain complete data or the full text of the original source information. They can be alphanumeric or textual in nature. Source databases are further sub-divided into 4 categories, as shown below:

2.1 Numeric

These databases contain original census or survey type data and/or statistically manipulated representations of data. The data are often in the form of time series. The database can also contain

headings, titles and explanatory notes. The best known Canadian
example is CANSIM (Canadian Socio-Economic Information Management
System).

2.2 Textual-Numeric (or Alphanumeric)

These databases consist of records that contain a number of data
elements or fields with a combination of textual information and
numeric data. Census or survey data are often in this form.

2.3 Properties

Such databases contain dictionary or handbook type data, typically
chemical and physical properties.

2.4 Full-Text

These data bases contain records of the complete text of some
primary source, such as a newspaper item, a specification or a
court decision. Such databases are becoming particularly important
in the fields of legal research/documentation and news/information
services.

4.2 Record Structures and Database Management Systems

Reference databases of both types are typically sequential files, with
records that contain a number of descriptive elements. Both types are

easily handled by today's information retrieval systems. Database
management capabilities need further upgrading to handle dynamic file
updating and modification, but generally available retrieval and
display capabilities are adequate and well suited to the majority of
intended users of such databases. Although users such as reference
librarians can be expected to have received some training in the use of
a particular vendor's system/package, ease of use remains an important
consideration.

The file and record structures of source databases are much more varied
than those of reference databases. Many numeric databases consist of
large numbers of time series, which are retrieved singly or in groups,
and over a specified time horizon. On the other hand, full text
databases and some of the textual-numeric databases may be more like
bibliographic databases in record and file structures.

One can therefore see the requirement for three types of database
management and retrieval system, each optimized for its particular
role: systems for handling bibliographic databases; systems for
handling numeric databases, especially those containing time series,
which require special manipulation; and systems for handling full-text
databases, which require a full-text, free-text search capability.
Each may be offered by a different type of vendor, and there may or
may not be economies of scope for the vendor attempting to offer two or
more different systems.

4.3 <u>Database Producers</u>

For purposes of analysis, database producers can be divided into two broad categories: those who produce reference databases and those who produce source databases.

In many cases, reference database producers are primarily publishers of printed index and abstract journals such as Chemical Abstracts. These organizations, which can be in either the private or public sector, acquire, screen, select, index and often abstract or summarize the primary literature. To produce printed publications efficiently, these organizations have adopted computerized systems for photocomposition and phototypesetting. This operation generates magnetic tapes that can be reformatted and used as direct inputs to computerized storage and retrieval systems.

Source databases, on the other hand, are produced by a number of different types of organizations. Some are government agencies that have a responsibility for the dissemination of the information collected or generated in their particular area; examples would include Statistics Canada and the U.S. Bureau of Labour Statistics for source numeric databases. Some producers are also publishers of printed reports, books, journals and newspapers, who have turned to electronic publishing to protect their revenues or increase their market share; examples would include newspapers (<u>New York Times, Globe and Mail</u>) and legal publishing firms (QL Systems Ltd. and West Publishing Co.) for source full text databases. Some producers process and package data that were collected by some other source, often a government agency. In their packaging, these producers frequently bring together data from a number of sources, and sometimes increase the value of a collection by including additional derived

data, such as summary totals, forecasts, etc. that they generate from the source data.

Most producers make their databases available to database distributors or vendors, under a variety of contractual arrangements. Some, however, ditribute their own databases. Such organizations, who combine the database production and distribution functions, are sometimes referred to as integrated services. They are described further in the next section.

4.4 Database Vendors

Database vendors can be classified into three groups: Remote Computing Services (RCS), Search Services and Integrated Services. Each group has special characteristics, which are described below. Appendix 1 gives a list of the database vendor firms who have their headquarters in Canada.

4.4.1 Remote Computing Services (RCS)

Timesharing service bureaus or remote computing services, sometimes referred to as "network information services", form the numerically largest category of database vendors. These firms provide a wide variety of data processing related service, including computing power, application and custom software development, EDP consultancy and educational services. Public and private-line telecommunications networks are used to deliver computing power as well as information to many users simultaneously, at locations remote from the cental computer site. A companion report of the Task Force has studied the Canadian commercial service bureau industry.

As a byproduct of providing interactive computing power, more than 80 RCSs also distribute online databases. This activity is either based on a deliberate marketing strategy, or is in response to customer requests; it generally remains a secondary activity and source of revenue. RCSs usually distribute source numeric or alphanumeric databases, where the data is often manipulated or post-processed by the user after retrieval, thus generating further revenues based on computer usage. Appendix 4 shows that 62 out of the 66 database distributed by I.P. Sharp Associates and all 57 of the databases distributed by DRI are source numeric in format. In such cases it is difficult, if not impossible to separate the revenues attributable to information retrieval from those related to post-processing of the retrieved data.

4.4.2 Search Services

The term "search services" is usually applied to vendors of reference/bibliographic databases. It refers specifically to their software programs and the types of user-computer interactions that they make possible.

The largest bibliographic search services in North America are Dialog Information Systems, Systems Development Corporation (SDC), Bibliographic Retrieval Services Inc. (BRS), the National Library of Medicine (NLM) and the Canada Institute for Scientific and Technical Information (CISTI). These services differ from other online vendors in that they have traditionally concentrated on distributing large numbers of reference/bibliographic databases. This emphasis may, however, be changing due to technological developments. The search services are refining their software to allow access to and manipulation of alphanumeric databases. IRD predicts that by 1985 the three major U.S. search services in the private sector, Dialog, SDC and BRS, will transform

themselves into full-fledged "all-purpose" distributors of the differing types
of databases: reference/bibliographic, source/alphanumeric and source/full
text.

4.4.3 Integrated Services

Organizations that both produce and distribute online databases are
referred to as "integrated services". This category overlaps, to some extent,
with the RCSs, the difference being more one of emphasis. With RCSs, databases
usually play a supporting role to the timesharing function; by contrast, for
integrated services both the databases and the distribution system play equally
important roles.

By the above definition, both I.P. Sharp Associates and DRI should,
perhaps, be classified as integrated services. Both firms produce a significant
number of the databases which they distribute. Many "electronic publishers"
provide integrated services. They include Dow Jones News/Retrieval, Mead Data
Central's LEXIS and NEXIS Services, the New York Times Information Service and
Info Globe. There has been a tendency for "electronic publishers" to acquire
established RCSs, and use them to disseminate a wide variety of databases. The
best examples are McGraw-Hill's acquisition of DRI (and its integration into the
McGraw-Hill Information Services Company), and the acquisition of The Source by
Reader's Digest.

5. <u>Market Structure and Characteristics</u>

In some respects, the market for public online information retrieval services is a worldwide one. It is not homogeneous, however, and it can be segmented into sub-markets, both geographically and by the type of online databases offerred. No sub-market, either geographical or functional, is monopolized by a single company. Each sub-market does, however, have a few dominant firms. Barriers to entry exist in both database production and distribution and both functions seem to be characterized by at least some economies of scale and scope.

5.1 <u>Total Market Size and Geographical Segmentation</u>

Although there are no reliable statistics, the world market for information retrieval and anciliary services was valued at U.S. $2 billion in 1979 (Ref. A15). An Input study (Ref. C5) gives a value of U.S. $1.17 billion for the domestic U.S. online database services market in 1979, and a value of $150 million for the corresponding European market. To see this market in perspective, it may be mentioned that $2 billion in 1979 was just over 1% of the world electronics market and under 5% of the computer market. Compared to the $12.5 billion turnover of the U.S. publishing industry in 1979, the online database services market of $1.17 billion was only some 9.4%. It should be stressed, however, that this market is growing much more rapidly than the publishing industry, or the computer market as a whole.

Because of the supporting telecommunications infrastructure and institutional factors, it is both possible and convenient to segment the

worldwide market into geographically distinct sub-markets such as North America (the U.S. and Canada), Europe (the EEC countries) and the Far East (Japan, Australia, Hong Kong, Singapore, etc.). The North American sub-market is served by an integrated, packet-switched data communications network which is steadily becoming universal in scope. A similar development may occur in the European sub-market by 1985. Although public and private telecommunications networks (such as Telenet, Tymnet and the Sharp APL Network) link these three sub-markets, a database vendor with computer facilities located in North America is still faced with substantial telecommunications costs in serving users located in the other two sub-markets and vice versa.

On the demand side, U.S. domestic demand is predominant and currently accounts for at least 80% of the total demand. Canadian demand is some 5% of U.S. demand, and European demand only some 15%. The European share is surprisingly low, given the fact that the total Gross National Product (GNP) of the EEC countries is higher than that of the U.S. Online database services made a slow start in Europe, and demand for such services is only now taking off.

Supply side information regarding database producers and vendors, as well as the number of databases and the volume of information they contain, is given in Tables 5.1 and 5.2. Table 5.1 shows that of the 512 database-producers listed in the Cuadra directory, only 61% are U.S., while 10% are Canadian and 25% belong to other countries. Table 5.2, which deals primarily with reference and textual databases, shows that while 59% of such databases were produced in the U.S. in 1975, 51% were produced outside the U.S. in 1979. The U.S. databases accounted for 88.5% of the total volume of records in 1975, and 63.2% in 1979. It should be noted, however, that a very large proportion of the non-U.S. database producers are government agencies or non-profit organizations.

Information regarding database vendors is given in Table 5.1 and Appendix 1. Of the 170 vendors listed in the Cuadra Directory, 62% were headquartered in the U.S., 11% in Canada and 27% in other countries. It should be noted, however, that while most vendors only serve their own national or regional sub-market, the major private North American vendors (all U.S. except for I.P. Sharp Associates Ltd.) also serve the European and Far Eastern sub-markets, either directly through subsidiaries or by joint ventures with local firms. There are virtually no corresponding examples of European vendors serving the North American sub-market. Many of the major U.S. vendors are also the dominant firms in their product sub-market. It is threfore quite possible that the U.S. share of total vendor revenues may be as much as 85-90%, rather than the 62% share indicated by the mere number of vendors.

5.2 Market Segmentation by Product Lines

Based on product differentiation, the market for public online information retrieval services can be segmented into three broad sub-markets. These will be classified as the Reference database, Source Alphanumeric (i.e. numeric and textual-numeric) database and Source Full Text database sub-markets. Table 5.3 gives the number of databases in each area, broken down further by four major disciplines.

Reference databases are used primarily by reference librarians, information specialists and research workers. Although accounting for some 40% of the total number of databases, they produce only some 10% of the vendor revenues. Less than 10% of such databases relate to Business and Economics, while over 50% relate to Science and Technology and some 20-25% are Multidisciplinary. As indicated in Section 4, the dominant vendors in this

sub-market are the major U.S. search services, especially Dialog, SDC, BRS, OCLC Inc. and the National Library of Medicine.

Source alphanumeric databases are heavily used by business firms and other ogranizations to aid managerial decision making and corporate planning. They can be further sub-divided into Financial, Socio-Economic, Demographic, Credit and Marketing databases. This sub-market contains some 50-55% of the total number of databases, but accounts for some 70-80% of total vendor revenues. Over 90% of these databases are related to Business and Economics. This sub-market is growing more rapidly than the Reference database one.

The Source Full Text sub-market may be sub-divided further into Legal and News databases. Although vendor revenues are currently small, some 5-10% of the total, they are growing rapidly. If the price/performance of news database services is sufficiently improved over the next 5-10 years by technological developments, a mass consumer demand may develop, with millions of subscribers rather than a few thousand information specialists and researchers. The other distinct possibility is a merger of the Reference and Source Full Text sub-markets, due to the possibility of storing the full text of a paper or article with its citation, rather than just a summary or abstract. The major search services can be expected to lead this development.

5.3 Barriers to Entry

Although the provision of and access to telecommunications facilities/services is heavily regulated in most countries, the public online information retrieval services industry is not. Its industrial structure shows

much less concentration than is found in telecommunications services or mainframe computer manufacture; there is no AT&T or IBM in the industry. However, three kinds of barriers to entry do exist:

   i) Investment and Capital requirements;

   ii) Access to relevant data;

   iii) Economies of scale (and scope).

5.3.1 Investment and Capital Requirements

  Although definitive information is not available, it seems clear that entering the online information retrieval services market require substantial initial investments relative to the revenues likely to be achieved in the first 3-5 years. For example, the New York Times invested about \$10m in 1973 to put its INFOBANK news database online; Mead Corporation's initial investment in its LEXIS legal database, which opened for service in 1973, was \$20m; TRW's investment was of the order of \$70m (Ref. A15).

  These investments are required to acquire a host computer and its operational/programming support staff, set up a telecommunications network, acquire or develop the necessary database management software, acquire a set of databases and convert them into the form required by the management/retrieval software and then load them online so that they can be accessed by end users who have been educated/trained to use the service. It may take the new entrant 2-3 years to accomplish the above process. A further period of time must elapse before the break even point is reached, when revenues cover operating costs.

Acquisition of an existing company (database vendor or timesharing computer service bureau) is, of course, an alternative form of entry investment. Thus McGraw-Hill acquired DRI for $103m. Similarly, Dun and Bradstreet acquired National CSS Inc., and Reader's Digest acquired Source Telecomputing Corporation.

For the database producer, there is a major difference between the investment necessary to set up a data collection mechanism and that necessary to put an existing machine-readable database online. An information system can be very expensive to set up. Producers questioned by LINK estimated that the investment needed to start from scratch could range from $2m to $10m. This is why many information producers are government agencies who have a mandate to create, collect or otherwise provide data; or print publishers who use computerized techniques to create input data for hard copy publications. Once the costs of information provision have been absorbed, information production becomes a marginal cost.

5.3.2 Lack of Access to Data

In addition to requiring large investments, access to relevant data for the purpose of producing an online database may be faced with other obstacles. The contractual permission of the creators/originators of the data is required if the data is not in the public domain. This permission may be witheld by authors and/or publishers for both economic and non-economic reasons. The definition of the property/economic rights of the information provider is still confused, to say the least. Agreements such as the Berne Convention and the Universal Copyright Convention have been made somewhat obsolete by the new information technologies. These legal/institutional issues are discussed further in Section 15.

5.3.3 Economies of Scale and Scope

As shown in section 5.3.1, database distribution involves substantial overhead fixed costs. On the other hand, the reproduction cost of a given item of information, excluding the physical medium, is virtually nil. Therefore, in an information retrieval system, the reproduction cost of an item should approach the marginal cost of accessing and retrieving it. Such a situation should give rise to important economies of scale.

One would also expect an inherent economy of scope between database distribution and the traditional activities of computer timesharing service bureaus. The computer facilities, support staff, telecommunications networks, user training and support are required for service bureau operations. The additional costs involved in database distribution are related to acquisition of databases and appropriate software, extra online storage, loading and updating the databases regularly and marketing/training costs. Source numeric databases present an additional attraction, because the information retrieved is usually manipulated or post-processed by the user, thus resulting in additional computer usage revenues. These characteristics are well demonstrated by the operations of I.P. Sharp Associates and DRI.

There could be further economies of scope between the activities of information providers and producers, and also between database production and distribution.

Table 5.1: Geographical Distribution of

Database Producers and Vendors

Area	Database Vendors		Database Producers		Databases Produced		Comments
	Number	% Total	Number	% Total	Number	% Total	
United States	105	61.8	314	61.3			
Canada	18	10.6	52	10.2			
Other Countries	47	27.6	130	25.4			
International Agencies (UN, OECD, EEC)	-		16	3.1			
Total	170		512				

Source: "Directory of Online Databases", Cuadra Associates, Vol. 3, No. 1,

published Fall 9181.

Table 5.2: Geographical Distribution of Reference and Natural

Language Databases and Database Records 1975–1979

Area	Item	Year						% Increase
		1975		1977		1979		1979 1975
		No. of Databases	No. of Records (m)	No. of Databases	No. of Records (m)	No. of Databases	No. of Records (m)	
United States	Data-bases	177		208		259		46.3
	% Total	(58.8)		(57.5)		(49.1)		
	Records (m)		46		58		93.5	103.3
	% Total		(88.5)		(81.7)		(63.2)	
Non U.S.	Data-bases	124		154		269		116.9
	% Total	(41.2)		(42.5)		(50.9)		
	Records (m)		6		13		54.5	808.3
	% Total		(11.5)		(18.3)		(36.8)	
Total	Data-bases	301		362		528		75.4
	Records (m)		52		71		148	184.6

Source: Williams, M.E. "Databases and Online Statistics for 1979" (Ref. B14).

Table 5.3: Analysis of Databases by Discipline and Type

Major Discipline	Reference Databases			Source Database				
	Biblio-graphic	Referral	Total	Numeric	Textual-Numeric	Properties	Full Text	Total
Business and Economics	12 (9.9%)	1	13 (8.8%) (6.2%)	179 (93.2%)	11 (91.7%)	–	5	195 (93.8%)
Science and Technology	72 (59.5%)	7	79 (53.7%) (85.9%)	3 (1.6%)	1	9	–	13 (14.1%)
Social Sciences and Humanities	19 (15.7%)	2	21 (14.3%) (67.7%)	10 (5.2%)	–	–	–	10 (33.3%)
Multidisciplinary	18 (14.9%)	16	34 (23.1%) (73.9%)	–	–	–	12	12 (26.1%)
Total	121 (82.3%)	26 (17.7%)	147 (39%)	192 (83.5%)	12 (5.2%)	9 (3.9%)	17 (7.4%)	230 (61%)

Source: "Directory of Online Databases", Cuadra Associates, Vol. 3, No. 1, published

Fall 1981.

6. Technological Trends

A technology profile report (Ref. A16) has been prepared as a stand-alone document for the Task Force. This report provides a general overview of the major areas and advances in computer technology and capability, emphasizing the present environment and changes envisaged over the next five years. The term computer technology is defined in the report to include hardware, software and telecommunications aspects of computing.

In this section, we shall limit ourselves to looking at trends in computer-communications technology which are relevant to the provision of public online information services. Before this is done, however, we will review developments made through 1955-80 and trace the evolution from batch searching to online search services provided to an unrestricted number of users, by information utilities operating on a national or multinational scale.

6.1 Historical Evolution of Online Information Services

The feasibility of applying a computer to bibliographic searching was first demonstrated in 1954 (Ref A6). This was followed by research and development efforts which, over the next ten years, led to a number of special libraries offering regular batch search services. In 1964, the U.S. National Library of Medicine (NLM) started a batch search service using its Medical Literature Analysis and Retrieval System (MEDLARS). This was the first large-scale, computer-based, retrospective search service to become available to the general public. Batch searching was a useful advance over manual searching for a number of questions, but its shortcomings precluded widespread applica-

tion. The average NLM in-house turnaround time was about two weeks, and the access retrieval timespan for the user was about six weeks. Online searching therefore developed as a viable alternative, being made technically feasible by the integration of computing technology with telecommunications networks.

Online systems began in a batch searching context and proceeded to evolve in three phases, as follows:

1) Feasibility Studies and Demonstrations

This phase lasted from 1960, when Systems Development Corporation (SDC) made the first public demonstration of online bibliographic searching, to about 1967. By the latter date SDC, Lockheed/Dialog, Stanford University, MIT and IBM had all developed and demonstrated online systems. SDC, supported by the Advanced Research Projects Agency (ARPA), had carried out an experiment in 1965 to demonstrate an online retrieval network on a national scale.

2) Production Services with Restricted User Populations

This phase lasted from about 1967 to 1971. The first efforts beyond the demonstration stage to regular production search services were made in 1967 by Lockheed. Regular online search services were provided to the NASA Ames Laboratory using the DIALOG System and a file of 260,000 citations. This was followed by other Lockheed services, SDC's first ORBIT-based services, online searching of NLM's MEDLARS databases using the IBM/STAIRS and AIM/TWX retrieval systems, and Stanford University's

University's SPIRES system. In each case, the number of users was limited, varying from 6 to 32, and access was via the public switched telephone network, or private leased lines. The number of users was also limited by the fact that at that time only a restricted number of interactive terminals could be supported by a computer at any one time.

A similar time frame witnessed the start and growth of computer service bureaus offering a variety of remote computing services, ranging from remote batch, through conversational remote batch to interactive time sharing services. Two companies in particular, I.P. Sharp Associates Ltd. and General Electric Information Services Company (GEISCO) (See Appendix 4), adopted the strategy of providing interactive services to an international market right from the beginning. The use of key driven terminals, especially CRTs, as input-output devices also began to increase. Some of the service bureaus began to offer numeric data bases to their clients either as a result of specific user demands or as a natural extension of their service offerings.

3) National or Multinational Information Utilities

The decade 1971-81 has witnessed a steady maturing, followed by a rapid growth of public online information retrieval services. From serving limited constituencies, they became available to an ever widening audience, then to anyone in North America who wanted to use the services, and finally (at least in theory) to anyone in the world.

The Lockheed DIALOG system opened its service to interested organizations in 1971-72 and began to expand its databases. Today it offers access to some 130 databases and serves over 13,000 users in 40 countries. The GEISCO service provides local calling access to its telecommunications network in some 700 cities in 25 countries. I.P. Sharp provides similar local calling access to its Sharp APL Network in some 400 cities in North America, Europe, Asia and Australia.

The removal of high long-distance calling charges completely altered the economics of using public information services. The key breakthrough in the 1970's, which enabled these database vendors to reach large audiences, was in the implementation of telecommunications networks designed specifically for data transmission; this has greatly improved the reliability, as well as the price/performance of the telecommunications links. Both digital data networks like Dataroute, and packet switched networks like Tymnet, Telenet, Datapac and I.P. Sharp's private network, have played an important role. Today all major vendors in the U.S. can be accessed via the U.S. international value added networks Tymnet and/or Telenet, and all major ones in Canada via Datapac (see Appendix 1). Furthermore, because of gateways between Telenet and Tymnet on the one hand, and Datapac on the other, the whole of North America is, for practical purposes, served by an integrated packet switched network. The private Sharp APL Network also provides gateways to these three packet switched networks, as well as to the Telex network. Progress in providing access from the North American packet switched networks to the European ones has been slower; many links from North America to Europe still use leased private

circuits. The North American vendor providing online services in Europe is still faced with large telecommunications costs. Further current policies of Europe's Euronet DIANNE network do not provide a hospitable climate for North American data base vendors.

6.2 Technological Trends in the 1980s

As stated earlier, the key technological components of public online information retrieval services are:

1) Powerful timesharing computers operating as nodes of a teleprocessing system which permit a large number of users to carry on simultaneous interactions with the system via remote terminals;

2) Large amounts of low-cost, random access storage media for holding the databases online;

3) Appropriate systems and applications software, capable of efficiently organizing large databases, flexible searching and retrieval, and having user-friendly interfaces;

4) An extensive, low-cost data transmission network capable of linking a wide variety of terminals to the central computers of the vendors;

5) Sophisticated user terminals which provide input-output capabilities, a communications interface and, preferably, some local intelligence.

Significant price/performance improvements can be expected to continue for those components whose performance is directly enhanced by Very Large Scale Integration (VLSI) micro-electronics technology. This is particularly true of the large central computers and the user terminals (Ref. A16). Some sources have claimed that the price/performance of core memory and processing power are increasing by 40% and 25% per annum. Although the most spectacular advances have been in the field of micro-computers, the large central mainframe computer has also benefited from these developments. User terminals, in addition to having higher transmission speeds (1200 baud and higher) and greater local intelligence, may also evolve into multifunction work stations. The use of such workstations, which would combine word processing and data processing functions with communications capabilities and stand alone use, is likely to become widespread by 1990. In some organizations, every professional worker and manager may be able to access a public online database from a desktop workstation.

Large amounts of random access storage capacity, provided principally by electro-mechanical disk drives, are required for major online systems. Online storage is needed for:

a) the basic operating system and search service programs; -

b) storage of master files;

c) storage of inverted indices to the files, which may have two to three times the storage requirements of the files themselves; and

d) storage of intermediate results from user searches.

As an example, consider the Lockheed DIALOG system, which provides access to some 130 online databases containing 40 million citations. The computer configuration currently operates with over 150 disk drives. Approximately 20 billion characters of information, mostly inverted index files revised as a result of each update, must be written to disk each month merely to maintain the files.

From 1960 to 1980, the storage capacity of disk drives has improved by two orders of magnitude, while the cost of storing a million bytes of information has decreased by almost the same factor. The 300-600 Megabyte capacity disks are now in routine use and disks with a capacity exceeding 1 billion bytes (gigabyte) are on the horizon. The most striking innovation in disk drive technology in the 1970s was the IBM Winchester drive which led to major improvements in the price/performance of disks. It comes in $5\frac{1}{4}$in, 8in and 14in disk sizes. Storage Technology's STC 8650 drive, announced in 1978, can store 10 billion bytes of information on its 16 disks (625 megabytes per disk). The main question in the 1980's is whether this rate of progress can be maintained using electro-mechanical technology. Some sources maintain that the next big breakthrough in disk storage will be adaptation of optical video disks to data storage. If the problem of erasing and rewriting data on video disks can be solved, price/performance improvements of another two orders of-magnitude might result.

Telecommunications costs declined steadily throughout the 1970s, at an average annual rate of some 10%. The big breakthroughs came with digital data networks, packet switched networks and satellite communications. It can be expected that progress in telecommunications technology in the 1980s will lead

to public switched broadband networks, both terrestrial and satellite based, and that packet switched networks will become universal in scope. The main question for the vendors and users of public online information retrieval services is whether the savings accruing from this technological progress will be passed on to them, in the form of improved price/performance, by the telecommunications carriers. Under regulatory regimes favouring monopoly supply, there is always the possibility that rates for data communications services may be kept artificially high, and the excess revenues used to subsidize other services.

The main uncertainty lies in the area of systems and applications software development. Improvements in hardware price/performance have not been matched by corresponding software improvements during the 1970s. Although reliability has increased, progress in improving the efficiency of organization, searching and retrieval techniques has been incremental. There may be a major breakthrough, however, in applications software for full text databases. The concept of the "database computer" also holds promise.

C. The Economics of Public Information Systems

This report highlights the lack of hard data regarding the demand side (usage volumes and revenues) of this industry generally, and the Canadian marketplace in particular. In this section, we present some preliminary findings regarding the U.S. and Canadian marketplaces for Public Online Information Retrieval Services, and attempt to identify the sub-markets in which trade in such services is likely. We have also examined the pricing structures adopted by vendors in different sub-markets, and attempted to identify trends.

The preliminary findings have been grouped into three broad areas: the U.S. marketplace, the Canadian marketplace and likely areas of trade in Public Online Information Retrieval Services. Many of the present findings are tentative, and may have to be modified as further information becomes available. The observations regarding flows of, and trade in services should be considered as speculative, since they cannot as yet be backed up by quantitative data.

7. Pricing Structures and Trends

Pricing policies for access to and use of online database services can be quite complex. There are a number of components to the prices and they can be combined in a number of different ways. There is also a basic difference in pricing practices between the vendors of numeric databases, and those of textual databases (Reference or Source Full Text). In the former case, post-processing of the information accessed is the rule, rather than the exception; the user wishes to display, manipulate or analyze the information accessed, or use it in conjunction with a forecasting model. Numeric database vendors, therefore make their money by charging for computer time used. Textual data, on the other hand, cannot be post-processed in the same way. Search services, therefore, tend to charge by connect time or search costs.

There are some general points which can be made about pricing and prices. First, many vendors (30% according to the Cuadra Directory) require some kind of subscription charge for providing access to their databases. These subscription charges range from less than a hundred to several thousand dollars annually. DRI, which pioneered this idea, charges $16,000 for its U.S. macro-economic service and $26,700 for its agriculture service. In some cases, the subscription fee entitles the user to a package which may include access to one or more databases, training on how to use the software, and a specified amount of consulting services. In other cases, the user may have several options, each consisting of a front-end subscription price and associated usage charges. Many vendors also give volume and commitment discounts on their standard schedule of usage charges. Such discounts are of particular importance

to large firms who expect to make heavy usage of a particular vendor's services. They serve to considerably reduce the marginal cost of using the service.

The two major pricing strategies, one used by the timesharing computer service bureaus for their numeric databases, and the other used by the rest of the vendors for textual databases, are discussed below. It must be remembered, however, that there are exceptions in each group.

7.1 Pricing by Timesharing Computer Service Bureaus

Most service bureaus impose a monthly minimum charge, ranging from $50-100, that is applied if the total usage charge for the given month does not reach the minimum level. The usage charges include at least the following components:

i) Connect Time

An hourly rate, ranging from $1.00 to $21.00, that is charged for the period during which the user is connected to the system. This figure may or may not include telecommunications costs. The use of a packet switched network (e.g. I.P. Sharp) may result in a very low connect cost combined with a cost for every packet transmitted. The connect rate may or may not vary with the transmission speed of the terminal being used. A 1200 baud terminal could be charged a higher rate than a 300 baud one.

ii) <u>Computer Resource Units (CRUs)</u>

A rate for actual usage of the system resources, which combines charges for use of the central processing unit (CPU), the amount of input/ output performed between the core memory and secondary storage (online disks and tapes), and sometimes charges for the user's work area in the core memory. The charges for CRUs vary enormously from vendor to vendor. They cannot be compared across vendors because of variations in the processing power of different vendor computers, and because the formulae used to calculate a CRU varies from vendor to vendor.

iii) <u>Online Storage Costs</u>

Additional costs are incurred by the user if he elects to store a selected amount of data from a public database, together with his own data, in a private online file. These costs are usually quoted in terms of daily or monthly rates per "page" or megabyte (million characters) stored. Service bureaus often provide large commitment discounts to users who agree to buy a large amount of online storage over a fixed period of time, such as a year. Such plans serve to reduce the marginal cost of storage.

iv) <u>Offline Printing</u>

The service bureau charges for offline printing on a high speed line printer. The rates range from $2-$5 per thousand lines printed. For large volumes of output, this is usually a more economical alternative than retrieving and printing the output at a key driven terminal because it avoids additional connect costs.

The online database customer may or may not be using the service bureau's standard schedule of charges, which usually offers discounts of varying magnitudes for non-prime time processing. In particular, the Computer Resource Units charged for the use of a particular database may be greater than the standard rate charged for other data processing services. This difference can occur because a surcharge has been added to the standard rates (by an additive or multiplicative factor) to act as a royalty to the producer of the database, or because the database system being used is more demanding of system resources than the "normal" mix of jobs. Lastly, the vendor may choose to levy a subscription fee or surcharge as a form of value pricing for access to and/or use of a particular database. This practice is used extensively by DRI, as can be seen in Table 4.2, Appendix 4.

7.2 Pricing by Search Services and Other Vendors

In most other cases, especially the use of reference bibliographic and source full text databases, costs are based on the following elements:

i) Connect Time

An hourly rate, which may or may not include telecommunications costs for network access, and royalty charges. In many cases all three costs are bundled together into one flat hourly connect time rate. In others the royalty cost is stated separately or the telecommunications costs are made the direct responsibility of the user. The range of connect time rates varies from about $25 to $300 per hour. Business databases tend to have higher rates than those in the physical or social

sciences. The Cuadra directory quotes an "average" rate of U.S. $65 per hour. Our preliminary review of the use of online information retrieval service in the federal government yields an "average" rate of Cdn. $70 per hour.

ii) Printing Costs

For bibliographic and some referral databases, there is an additional charge for offline printing, which is generally based upon the number of citations or pages printed. This charge may range from a low of $.05 to a high of $5.00 per citation (the average is about $0.15). There may also be charges for online printing of the retrieved information at the user's terminal.

With the bibliographic databases, users can estimate costs across various online vendors by using the concept of a "search". Cuadra states that a search of a single database generally takes less than 15 minutes and may cost about $25 on the average. Averages, however, are just that. A simple search may require only 2-3 minutes, while a complicated one may take an hour. Lawton (Ref. A6) quotes "search" prices ranging from Cdn. $4 for health databases to Cdn. $43 for business databases.

One firm, QL Systems Limited, has built its pricing strategy around the concept of the "search", rather than connect time (See Appendix 2). It should be noted that there is no comparable "search" measurement for the use of numeric databases.

8. The U.S. Marketplace

8.1 Overall Size and Growth Trends

(1) Public Online Information Retrieval Services is already a billion-dollar industry in the U.S. and it is growing very rapidly. Order-of-magnitude estimates regarding the size of the industry are probably reliable, but data given in industry studies need to be checked for source and quality. Such verification has not been possible for the present report.

(2) Governments and non-profit agencies are major providers of information, but commercial enterprises predominate as information producers and database vendors.

(3) Available estimates of current industry revenues range from $500m in 1979 (Kiechel, Ref. A3), through $630m (CSI, Ref. C2) to $800m (Link, Ref. C4) in 1980, to $1250m in 1981 (IRD, Ref. C1). All these figures are in current U.S. dollars. The IRD estimate, however, includes over $150m of revenues from transactional databases and it also includes revenues to both database producers and vendors. Thus the estimates of revenues for 1980 to vendors of Public Online Information Retrieval Services lie in the range $600-900m, with an average value of $750m. Only Input (Ref. C5) comes up with a much higher estimate of $1.4 billion for the size of the 1980 U.S. Online Database markets. Further analysis is required to reconcile differences of this magnitude.

(4) High growth rates are projected over the next 5-10 years, but there is some difference between IRD and the other sources. IRD projects a quadrupling of revenues to $5.5 billion (in 1981 dollars) in 1991. This is equivalent to a real compound annual growth rate of 16% over the next 10 years. LINK, starting from a base of $795m in 1980, projects a nominal 34% compound annual growth rate over the next 5 years to $3.4 billion in 1985. CSI, starting from a $630m base, projects a nominal 38% compound annual growth rate to $3.0 billion in 1985. Input, starting from a much larger $1.4b base, projects a nominal 25% compound annual growth rate to $4.2b in 1985. Kiechel states that revenues are increasing at a rate "well over 30% a year". The difficulty in comparing IRD's projected real growth rate with the other nominal growth rates stems from the fact that no deflator of revenues is available. Depending upon the allowance made for inflationary increases over the next decade, IRD's estimates may be conservative compared to those of the others.

(5) The number of databases available to U.S. users, and the number of online service vendors offering these databases, are both growing rapidly. The fall 1981 issue of the Cuadra Directory (Ref. B3), lists 965 databases offered by 170 domestic and foreign vendors. The corresponding numbers listed in the spring 1981 issue of the same Directory were 770 and 135 respectively; in the fall 1980 issue they were 615 and 93 respectively. It may be noted here that European Association of Information Services (EUSIDIC) in their 1981 Database Guide state that over 1400 "databases" in every field of interest are now (1980) available to the European public (up from a little over 400

in 1975). Closer analysis reveals that only some 500 are accessible online, the others being offered only for batch or offline operations. In the context of this study, the latter would be considered to be data files.

(6) IRD's breakdown of the online database service revenues into sub-markets, and the growth projections by sub-market to 1991, are shown in Table 8.1. Input's breakdown for 1979 and 1985, and growth projections are shown in Table 8.2. LINK's breakdown of online distributor revenues in 1980 is shown in Table 8.3. All these sources highlight the relatively small size of the Reference/Bibliographic sub-market, and the crucial importance of the Source/Numeric sub-market, which includes Financial, Socio-Economic, Credit, Marketing and Demographic databases. According to IRD, this sub-market produced revenues of $937m (74.6% of the total) in 1981 and is forecast to produce revenues of $4,387m (78.6% of the total) in 1991. This is an increase of 368% over 10 years, equivalent to a compound annual growth rate of 16.7%.

(7) Revenues of the top 10 U.S. database vendors are shown in Table 8.4. These 10 firms had combined revenues of $255m, which is 34% of the average value of $750m estimated for the size of the U.S. market. Numbers of subscribers for the top 20 companies (ranked by subscriber count) are shown in Table 8.5. The total number of subscribers in October 1981 was over 150,000. Comparison of Tables 8.4 and 8.5 shows that there is no clear correlation between customer revenues and counts. Both McGraw-Hill and Mead Data Central typify vendors who have adopted the high price - low volume strategy in going for major

clients. Compuserve and The Source represent the low price - high volume model for consumer information services.

(8) By way of comparison, the revenues of the U.S. computer services industry were estimated by Input Ltd. as $14.9b in 1980, of which $8.8b came from processing services. A $750m market for public information services would be about 8.5% of the market for processing services. A size range of $600m (low) to $1.4b (high) would correspond to 6.8 - 15.9% of the market for processing services. In making these comparisons, the problem of potentially double counting the revenues of certain computer service bureaus must be kept in mind.

8.2 Reference Databases

Reference databases account for some 40% of the total number of online databases, but only some 10% of the revenues. The reference database sub-market in the U.S. is dominated by three major private search services and two non-profit organizations. These are:

i) Dialog Information Systems, a subsidiary of Lockheed Corporation, which offers access to some 130 databases via its DIALOG service. It had revenues of $25m in 1980, and some 13,500 subscribers in October 1981 (See Tables 8.4 and 8.5).

ii) Systems Development Corporation, (SDC), a subsidiary of Burroughs Corporation, which offers access to over 70 databases via its ORBIT service. SDC had some 6,000 subscribers in October 1981.

iii) <u>Bibliographic Retrieval Services</u>, (BRS), a subsidiary of Information Handling Services, which offers access to some 40 databases; it had some 2,000 customers in October 1981.

iv) <u>The National Library of Medicine</u>, (NLM), a non-profit organization which offers the Medline Service. Some 1,550 institutions currently subscribe to Medline. About 1.8m searches were recorded in 1980, and the 1981 total is expected to be even greater.

v) <u>OCLC Inc</u>., a non-profit organization whose primary customers are some 2,700 libraries. It had revenues of $27m in 1980. In October 1981 it had 2,700 subscribers accessing its services via some 4,100 terminals. Subscribers and terminals are growing at about 20 and 50 per month respectively.

IRD expects the "Scholarly research" sub-market, which covers the major portion of the reference database sub-market, to increase from $51.5m in 1981 to $235m (in constant 1981 dollars) in 1991. This is a 16% compound annual growth rate over 10 years, for a total growth of 356%. CSI forecasts a nominal 27% annual growth rate in reference database revenues, from $65m in 1980 to $220m in 1985. CSI estimates that some 3.5 million searches were conducted through U.S. public places in 1980. Some surveys indicate that only 5% of the potential market has been tapped.

Reference/Bibliographic databases tend to be less business oriented than source databases. Studies by LINK (Ref. C4) and Wanger et al (Ref. A5) have found that more than 50% of the bibliographic/referral databases deal with

the physical sciences and technology, while another 30% deal with the social sciences and humanities, or are multidisciplinary. Reference databases dealing with business subjects account for only some 10% of the total number.

Partly because of their non-business orientation, reference databases tend to be less expensive to use than source ones. Connect charges for the use of non-business reference databases range from $25-75 per hour. Since the price/cost margin is narrower, and the overhead fixed costs very considerable, the vendor must attract a large number of users to make a reasonable profit.

8.3 Source Numeric Databases

Source numeric databases constitute both the largest sub-market, and one of the fastest growing ones. These databases are often used by decision makers directly, without the intervention of intermediaries. IRD breaks down this sub-market further into three segments: Financial/Economic, Credit and Marketing/Demographic (See Table 8.1).

8.3.1 Financial/Economic

Revenues for this area are expected to increase by 352% from $342m in 1981 to $1,547m in 1991; the market share being about 27% of the total. These databases are used by business executives to help them make a wide variety of corporate decisions extending through all phases of corporate and financial planning.

Many of these databases are econometric. They are often marketed along with econometric forecasting models, model building software and economic consulting services. Data Resources Inc. (DRI), Chase Econometric Associates, Wharton Econometric Forecasting Associates and Citibank are the major producers of econometric databases. The first three organizations are subsidiaries, respectively, of McGraw-Hill Inc., Chase Manhattan Bank and the Ziff-Davis publishing company.

Financial analysts and stockbrokers make very extensive use of public information services. Access to and use of continuously updated stock market information has virtually become a condition of doing business. Penetration of this market by vendors like Quotron and the Dow Jones News/Retrieval Service is virtually 100% in all major cities, and quite high even in smaller ones. Recently, Dataline (Toronto) has begun offering its CANQUOTE service of updated stock market information.

8.3.2 Credit Databases

Revenues in this area are expected to increase by 368% from $450m in 1981 to $2,108m in 1991, and the market share from about 36% to 38% of the total. There are four types of credit databases: consumer credit, commercial credit, cheque evaluation and credit card verification. They are used by organizations, mostly banks and retail stores, for assessing the financial reliability of a customer or the validity of the potential customer's promise to pay.

i) <u>Consumer Credit Databases</u>

Over 80% of the 78 million U.S. households are currently logged on one or more consumer credit databases. The leading producer of such databases is TRW. Others include Equifax, Trans Union and ACS. The immense amounts of data to be gathered, its frequently unreliable nature, and the enactment of consumer protection legislation all combine to increase the costs of doing business for consumer credit database producers.

ii) <u>Commercial Credit Databases</u>

Commercial credit reports contain information on such matters as the payment habits of a company, the number of its employees, balance sheet information and the names and histories of its officers/directors. The leading producer of Commercial Credit Ratings is Dun and Bradstreet, whose Business Information Centre provides this information to the U.S. business community. The Dun and Bradstreet Business Information File holds current information on more than 4 million U.S. businesses. The Centre retrieves requested information and mails nearly 50,000 reports daily to subscribers. The file is also available for online searcing from remote locations. TRW is the other major company in the commercial credit business.

iii) <u>Check Evaluation Databases</u>

These databases provide subscribers, usually banks and large retailers with historical information about the reliability of the subject's

cheques. The data are received from participating financial institutions. The largest producer of such databases is Telecredit Inc.

iv) Credit Card Verification Databases

These databases perform two functions. First, they provide information on whether a credit card has been stolen or is otherwise invalid. Second, they tell the subscriber, usually a retail store, if the amount of the proposed credit purchase falls within the card's established credit limits.

Credit card verification is usually performed by the card issuers themselves, e.g., American Express, or the bank card associations like VISA and Mastercard. The revenues of credit rating services are directly linked to the demand for credit and thus responsive to national and regional economic trends. Revenues also depend in large measure upon prevalent payment procedures. Thus the gradual move from paper money and cheques to plastic money (credit cards) has provided a boost to the credit information suppliers. On the other hand, a widespread move to EFT may have the opposite effect. By setting up a direct transfer mechanism between the purchaser's and retailer's bank accounts, EFT takes the time lag out of financial transactions, the very time lag which is responsible for creating the need for credit evaluation.

8.3.3 Marketing and Demographic Research Databases

IRD projects significant growth of revenues in this area, from $145m in 1981 to $732m in 1991 which is an increase of 405%, equivalent to a compound annual growth rate of 17.5%. Certainly there is no lack of data for producers interested in developing databases or vendors selling that information to users studying marketing issues. The 1980 U.S. Census alone collected over 3 billion items of data. The private sector database producers and vendors are helped enormously by the fact that the Census Bureau's data can be obtained by them and used for very little cost. Thus the producers are spared almost all of the front-end costs of data gathering, verification and editing.

These databases actually consist of two sub-categories:

i)   databases which contain information about potential customers and the products they use; and

ii) databases which contain information about product sales and the media used to establish contact with those markets.

Significant producers of databases about potential product purchasers include Urban Decision Systems Inc., which provides customers with detailed demographic information about virtually any domestic geographic area, no matter how unusually contoured, Demographic Research Company with its Zip Code Demographic Database and CACI Inc., which produces the Site II and Site Potential databases.

A.C. Neilsen is the premier company among producers of databases about the advertising effectiveness of various media. In addition to producing the

Neilsen Retail Index, the Nielsen Station Index and the Neilsen Television Index it is also an online distributor. A new entry into the marketing database area is AMI, a joint venture of J.W. Thompson and the New York Times Information Bank; its activities are discussed further in the next section under the New York Times Information Service.

8.4 Source Full Text Databases

This sub-market can be sub-divided further into Legal Databases and News Databases. The full text approach has been well received within the legal community and large scale penetration of this market is only a question of time. The customer base for general news retrieval services is still small and open to some question. However, should full text access become acceptable to the general consumer population, online database service vendors will be able to count their customers in the millions, instead of the thousands.

8.4.1 Legal Databases

Large U.S. legal firms make extensive use of full-text legal databases. The dominant vendor in this area is the Mead Corporation with its LEXIS service, which had some 800 users in October 1981. LEXIS costs include a minimum $500 monthly charge per terminal, plus connect costs of $60-90 per hour. It is facing competition from a new entrant Westlaw, which is offerred by West Publishing Company. These vendors are now trying to penetrate the medium size and smaller legal firms.

The IRD report estimated that revenues from legal databases would increase from $55m in 1981 to $197 in 1991, an increase of 258% in constant terms. The Information and Database Publishing Report (IDP) suggests that this might be an underestimate, because it apparently discounts potentially significant markets such as in-house corporate legal work and governmental use at federal, state and local levels. These markets may be presently excluded to a large extent by current pricing rates. But if major advances are made in searching capabilities, combined with marked reductions in data storage costs, as seems likely with the new videodisk technology, these may combine to open up significant new markets in the field of legal research, with its constantly growing body of case law and the need to capture existing case law retrospectively.

8.4.2 News Databases

News databases are expected to be one of the fastest growing areas in the 1980s. The maturing of news information services within public and private organizations, which are beginning to come into their own, combined with the expected evolution in the home information market, should fuel the demand side. On the supply side, major advances in full text storage and retrieval technology are expected to improve price/performance to the extent that news databases could evolve from a research tool for business information specialists to an item of mass consumer interest. The trend from bibliographic formats to full text delivery can be expected to continue.

A number of U.S. vendors already provide full text delivery of news reports. They include the New York Times Information Bank, Dow Jones

News/Retrieval, Mead Data Central's Nexis Service, and CompuServe. These are
expected to be joined by a variety of new entrants. The IRD report predicts a
386% revenue growth from just $28m in 1981 to $136m in 1991. This is equivalent
to a 17.1% real compound annual growth rate over the decade 1981-91. Two of
these services are described briefly below.

a) The New York Times Information Service is a subsidiary of the New York
    Times Company. Online access to NYTIS is available, by subscription,
    in the U.S., Canada, Mexico and Europe. The service became
    commercially operational in 1974. It implemented an enhanced full text
    free text, easy-to-use search system in the summer of 1981. NYTIS
    provides access to three databases:

    1) The Information Bank, which holds nearly 2 million references to
        general and specialized publications, including the New York Times,
        Wall Street Journal, Washington Post, Business Week and Financial
        Times;

    2) Advertising and Marketing Intelligence (AMI), a bibliographic
        database providing current information about people, new products,
        consumer trends, promotional programs, media planning and- buying;

    3) Key Issues Tracking (KIT), a current affairs database created
        jointly by the Executive Office of the President, the Department of
        Transportation and the Department of Energy.

Usage costs with no commitments vary from $100 - $140 per connect hour, depending upon the database. These hourly connect rates can be substantially reduced under a number of optional commitment contracts.

b) The Dow Jones News/Retrieval Service provides online access to business, financial and economic news from international and domestic sources, to stock market price data and to money market changes. The service is available by subscription. A special Auto-Entry terminal, which was developed to facilitate access, can now be leased from Dow Jones.

News/Retrieval contains up to 90 days of news covering 6,000 publicly held companies. The Stock Quote Reporter enables subscribers to monitor trading quotations on more than 6,000 publicly held stocks. Input to the system comes in the form of up-to-the-minute information from the Dow Jones News Service, Disclosure, Wall Street Journal and Barrons. Quotations are available from the New York Stock Exchange, American Stock Exchange and the Boston, Philadelphia, Midwest and Pacific Exchanges.

Dow Jones News/Retrieval Service had estimated revenues of $18m in 1980. As of October 1981 it had more than 20,000 subscribers, and was adding terminals at the rate of 2,000 per month. A more recent estimate (IDP Vol. 2, No. 21) is 30,000 terminals as of January 1982. Users without commitments pay hourly connect charges of $60 during business hours, and only $12 during non-prime hours. Users paying a $50 per month subscription fee get reduced prime time rates.

8.5 Services to the Mass Consumer

Two companies currently provide public information services to the mass consumer on a commercial basis. They are Source Telecomputing Corporation, now owned by Reader's Digest, and CompuServe Inc., now owned by H&R Block. As of October 1981, The Source had 11,500 subscribers and CompuServe's Information Service had 15,000. Prices, while much cheaper than those charged by the major search services like Dialog, are still significant. The Source charges customers a one time sign-up fee of $100, plus a monthly minimum of $10, whether the system is used or not. Hourly connect rates vary from $4.25 to $30, depending upon the databases accessed and the time of day; substantial discounts are given for use during off-peak hours. CompuServe requires a smaller initial sign-up fee of $19.95 but higher hourly connect charges during business hours.

Both The Source and CompuServe are discovering that the fastest growing segment of their customers' demand is for business services. Responding to this demand, The Source has announced a "new series of electronic super services", dubbed The Source PLUS. The new services include Legi-Slate, a database which tracks Congressional legislation, COMPUSTAR, an electronic shopping service; Management Contents Ltd., which offers abstracts from 27 leading business publications, Commodity News Service and Media General. The Source Plus hourly connect rate for 300 baud terminals is $30 during prime time on weekdays, $15 for evening or weekend use and $10 after midnight. The corresponding rates for 1200 baud terminals are $40, $25 and $15 respectively.

CompuServe continues to maintain its position as the leading consumer oriented online information service in the U.S. Its number of subscribers

increased by some 3,000 (25%) from 12,000 in mid-summer to 15,000 in October 1981. CompuServe Inc., the parent company of CompuServe Information Services, had revenues of $27.6m in 1980, an increase of 41% over the 1979 revenues.

Both companies expect to have a major increase in the number of subscribers by 1985. Some executives at The Source have claimed that it alone could have 500,000 subscribers by 1985. The future of these two companies could be greatly affected by the success (or otherwise) of videotex services over the next 5 years. A LINK report (Ref. D11) estimates that by 1985 there will be up to 2.5 million homes in the U.S. receiving videotext or teletex services over telephone lines, cable or broadcast.

IRD estimates that revenues from consumer databases amounted to $3m in 1981. Its expectations are that databases for consumers will become an important market by 1991, with total revenues of $795m for both transactional and public information services. It is expected, however, that most of the growth in the home consumer sub-market will come after 1985. IRD's estimates for the home information services sub-market are $3m in 1981, $21m in 1983, $300m in 1986 and $795m in 1991.

Several reasons are given for this slow initial growth. First, consumers need to be educated about the scope and usefulness of online services. Second, advertisers will be reluctant to lend their full support to such services until a "critical mass" of subscribers is reached. Third, the need for coordination between a wide variety of suppliers, including television set and terminal manufacturers, personal computer manufacturers, telephone companies, cable companies and information providers/producers, may slow down progress.

Finally, it may take another 4-5 years before costs of production decline to a point that the savings, when passed on to consumers, will bring about a widespread penetration of the home information services sub-market. The role of videotex services, acting as a user-friendly "Gateway" to a wide variety of transactional and public information services, may become particularly important.

Table 8.1: U.S. Online Database Service Revenues 1981 - 1991

(Millions of 1981 U.S. $)

Database Service Markets	Type of Database		1981		1983			1986			1991		
	Primary Classification	Secondary Classification	Revenue ($m)	% Total	Revenue ($m)	% Total	% Growth from 1981	Revenue ($m)	% Total	% Growth from 1981	Revenue ($m)	% Total	% Growth from 1981
Scholarly	Reference	Bibliographic, Referral	51.5	4.10	72	3.67	40	117.5	4.00	128	235	4.21	356
Financial/Economic	Source	Numeric	342	27.22	468	23.88	37	769	26.16	125	1,547	27.71	352
Credit	Source	Numeric	450	35.81	858	43.78	91	1,195	40.65	166	2,108	37.76	368
Marketing and Demographic	Source	Numeric	145	11.54	198	10.10	37	334	11.36	130	732	13.1	405
Legal	Source	Full-Text	55	4.38	75	3.83	36	117	3.98	113	197	3.53	258
News	Source	Full-Text	28	2.23	36	1.84	29	59	2.01	111	136	2.44	386
Patents			7		10			14			26		
Government/Watch			10		14			25			64		
Lists			3		4			7			19		
- Sub-Total			1,091.5	86.87	1,735	88.52		2,637.5	89.73		5,064	90.70	
Financial/ Transactional			160		214		34	272		70	438		174
Other Transactional			5		11			30			81		
- Sub-Total			165	13.13	225	11.48	36	302	10.27	83	519	9.30	215
- TOTAL			1,256.5	100	1,960	100	56	2,939.5	100	134	5,583	100	344

Source: International Resource Development (IRD) Inc. (Ref. C1)

Table 8.2: U.S. Information Retrieval Market

(Millions of U.S. $)

Subject Matter/Area	Type of Database		1979		1985		Growth from 1979	
	Primary Classif.	Secondary Classif.	Revenue ($m)	% of Total	Revenue ($m)	% of Total	Total %	Comp. Ann. Rate (%)
Bibliographic	Ref.	Bibliog.	62	5.30	255	5.96	311	27
Econometric	Source	Numeric	67	5.74	200	4.68	199	20
Economic and Financial	Source	Numeric	125	10.70	400	9.36	220	24
Stock and Commodity Markets	Source	Numeric	140	12.0	300	7.02	114	13
Credit	Source	Numeric	240	20.54	724	16.94	202	20
Market Research	Source	Numeric	105	9.0	510	11.93	386	30
Demographic	Source	Numeric	36	3.08	135	3.16	275	24
Natural Resources			16	1.37	96	2.25	500	35
Industry Specific			37	3.16	126	2.95	241	23
Real Estate			45	3.85	175	4.09	289	25
Patents			22	1.88	95	2.22	332	28
Legal and Accounting	Source	Full-Text	75	6.42	360	8.42	380	30
News	Source	Full-Text	40	3.42	235	5.50	488	34
International	Ref.	Bibliog.	18	1.54	70	1.64	289	25
Other			140	12.0	594	13.88	324	26
TOTAL			1,168		4,275		266	24

Source: Input Ltd. "International Market Opportunities for on-line database services"
Sept. 1980 (Ref. C5).

Table 8.3:   Online Distributor Revenues for 1980

Interactive Public Databases by Discipline

(Revenues in Millions of U.S. $)

Discipline/ Subject matter	Type of Database					
	Source		Reference		Total	
	Amount ($m)	% Total	Amount ($m)	% Total	Amount ($m)	% Total
1. Scientific and Technical	14.04	1.91	18.15	29.20	32.19	4.05
2. Business and Social Science						
– Business			6.04	9.72		
– Social Science and Hum.	690.36	94.14	6.25	10.05	702.65	88.33
			12.29	19.77		
3. Multidisciplinary	28.94	3.95	31.72	51.03	60.66	7.62
4. TOTAL	733.34	100.	62.16	100.	795.50	100.
		92.19		7.81		100.

Source: LINK Resources (Ref. C4).

Table 8.4: Top Ten U.S. Database Distributors Ranked by

Estimated 1980 Revenues ($m)

Distributor	Revenues ($m)	% of Total	Remarks
McGraw–Hill/DRI/ Standard and Poor's	53	20.78	°Socio-economic, Financial and Credit Rating services.
Mead Data Central	34	13.34	°Provides legal (LEXIS) and news (NEXIS) full text databases.
OCLC	27	10.59	°Non-profit corporation providing library systems/services in a shared, cooperative mode.
Dialog (Lockheed)	25	9.80	°Largest Search Service; over 130 Reference databases, 40m citations.
Equifax, Inc.	25	9.80	°Consumer credit databases.
Bunker Ramo	20	7.84	°Banking services to domestic and international markets.
Reuters	18	7.06	°News databases, money market services.
Dow Jones and Co.	18	7.06	°Worldwide business, economic and financial news; stock market information.
Dun and Bradstreet	18	7.06	°Largest private source for commercial credit and business information.
PRC Realty	17	6.67	°
TOTAL	255	100.	

Table 8.5: Selected U.S. Database Distributors Ranked by

Estimated Customer Count

Distributor	Customer Count			1980 Revenues ($m)
	Mid-1981	Oct. 1981	3 month % change	
TRW Inc.	22,000	24,000	9	
Equifax Inc. (1)	22,000	(22,000)	—	25
Dow Jones and Co.	15,000	20,250	35	18
Control Data Corp.	15,000	15,000	0	
CompuServe	12,000	15,000	25	
Dialog (Lockheed)	13,000	13,500	4	25
The Source	11,000	11,500	4	
PRC Realty (1)	10,000	(10,000)	—	17
GE Info. Services	6,000	6,000	0	
Systems Development Corporation (SDC)	6,000	6,000	0	
Comshare Inc. (1)	4,500	(4,500)	—	
Telerate	3,500	4,000	14	
Bunker Ramo Corp.	3,500	3,500	0	20
Reuters/Monitor/Dealing	3,000	3,000	0	18
OCLC	2,621	2,700	3	27
Dun and Bradstreet (DUNSPRINT)	2,231	2,400	7	18
New York Times Information Service	2,000	2,300	15	
BRS, Inc. (1)	2,000	(2,000)	—	
Tymshare (2)	—	2,000	—	
Chase Econometric/IDC	2,000	2,000	0	
Medline (2)	—	1,550	—	
McGraw-Hill/DRI/S&P	1,000	1,150	15	53
Mead Data Central	750	800	7	34
Total: Excluding (2)	159,102	171,600	8.0	
Including (2)		175,150		
Excluding (1), (2)	120,602	136,650	10.4	—

Source: IDP Reports, Vol. 2, No. 13, August 14, 1981.
Vol. 2, No. 17, October 23, 1981.

Notes:

(1)   Companies for which October 1981 count not available; set equal to mid-1981 count.

(2)   Companies for which mid-1981 count not available.

9. The Canadian Marketplace

9.1 Overall Size and Growth Trends

The Canadian federal government and its agencies have taken a more active role in the production and distribution of online databases than is now the practice of their U.S. counterparts. Examples would include CANSIM, CISTI and Telidon (Appendix 2, 3 and 5).

It is difficult to attempt a precise estimate of the size of the Canadian marketplace, or its growth rate, since no published data are available. An order-of-magnitude estimate of the size in 1980 would be greater than $10m and less than $100m. It is inappropriate to assume that the Canadian market in 1980 was 10% of the U.S. one, which was estimated at $600 - $900m (U.S.). If an estimate of 5-7%, corresponding to a smaller proportionate size and a 1 year time lag is used, this would give a range of $20-45m, with an average value of $32.5m.

For lack of a better estimate, the real and nominal growth rates for the overall Canadian market will be assumed as 16% (IRD) and 30% respectively. This nominal growth rate would produce a 1985 market size range of $75-167m. On the other hand, if the high CSI estimate of a 38% compound annual growth to 1985 is used, this would produce a market size range of $100-$225m.

Evans Research Corporation has estimated (Ref. B12) that the 1981 Canadian services market for online public databases, including exports, will be about $30m. The report also states that "If more Canadian service companies,

and more importantly Canadian firms, learn the usefulness of public databases, the Canadian market could be $100m by 1985". Although the basis for these estimates is not explained, they are well within the range of the above estimates, the only problem being the potential magnitude of exports.

By way of comparison, Evans Research Corporation estimated the 1980 revenues of the Canadian computer services industry as $1,060m. The revenues of the top 35 Canadian computer service bureaus were $463m in 1980. CADAPSO estimated processing revenues to be 78% of the total, or $827m. A more realistic estimate of 1980 processing revenues, made by DOC, is $575m. Using the lower figure of $575m would give a public information services market of 3.5-7.8% of the processing services market; using the higher figure of $827m would give a market size of 2.4-5.4%. In section 8.1, it was shown that the comparable U.S. figure is 8.5%, with a range of 6.8-15.9%. Assuming that the Canadian market for public information services lags the U.S. one by a year, a range of 6-7.5% of the processing services market may be an appropriate assumption. This would produce a range estimate of $34.5-$43.0m.

9.2 Reference Databases

The general observations made about Reference databases in describing the U.S. sub-market (Section 8.2) are also valid for the Canadian sub-market. The main difference is that the three major U.S. private search services, Dialog, SDC and BRS, compete strongly in the Canadian sub-market with the Canadian vendors, CISTI, QL Systems Limited and Informatech. The operations of CISTI and QL Systems are described in detail in Appendix 2. SDC's services are marketed in Canada by Infomart. The Health Sciences Resource Centre of CISTI is

Canada's national coordinator for access to bio-medical information in the MEDLAR'S databases of the U.S. National Library of Medicine.

Most Canadian online search service centres are located in governmental and academic organizations, more specifically, in libraries or information centres. Many Canadian reference libraries, including about one third of the libraries in post-secondary institutions, offer online bibliographic search services. Most of these libraries subscribe to CISTI's CAN/OLE System or QL/Search, as well as one of the U.S. search services. Recent studies (Ref. A6, A14) have shown that the average reference library subscribes to three search services. SDC/ORBIT, CAN/OLE, Dialog and QL Systems represents the ranking of search services by the number of centres using them.

Many Canadian online search service centres (45.5% according to Ref. A14) operate on a free basis for their customers. Each centre conducted an average of 430 online searches per year in 1978-79 while the total number of searches was 161,280 for the 380 centres surveyed.

CISTI and QL Systems Ltd. had revenues of about $2m in 1980. Informatech's total revenues were about $0.5m. Assuming that these three Canadian vendors accounted for some 40% of total use, this would produce an estimate of $6m for the size of the Canadian sub-market in 1980.

9.3 Source Numeric Databases

CANSIM is the most widely distributed and used socio-economic database in Canada, with one primary and ten secondary distributors, including most of

the important service bureaus. Precise usage statistics are not available as
yet, but there were approximately 4m series retrievals in 1980 (See
Appendix 3).

I.P. Sharp Associates is the largest Canadian vendor of source numeric
databases. It offers over 60 databases (see Appendix 4) and 20 million time
series in such areas as Economics, Finance, Aviation, Energy and Insurance, as
well as a worldwide packet-switched telecommunications network providing access
from over 400 cities in North America, Europe, the Far East and Australia. In
1980-81, the firm had total revenues of $35.5m, but it is difficult to assign a
specific figure for information retrieval. About one-third of Sharp's revenues
came from domestic Canadian usage, another one-third from exports to the U.S.,
and the remaining one-third from exports to Europe and the Far East.

No firm estimates can be made at this point regarding the size of the
sub-market for numeric databases, which includes socio-economic, financial,
credit, marketing and demographic databases. If the U.S. experience is any
guide, this is by far the largest sub-market and could constitute up to 75% of
the total market for public information services. A 75% market share would give
a sub-market size range of $15-34m. It is possible, however, that the Canadian
sub-market is developing more slowly than the U.S. one, because Canadian firms
are making less use of such services to aid the decision making process.

9.4 Source Full Text Databases

There are two Canadian vendors in this area: QL Systems and Info Globe.
Their activities are described in detail in Appendix 2.

QL Systems is the only Canadian vendor providing legal full-text databases. Current usage is very small, but growth prospects are considered excellent in the 1980s. Potential clients include the courts and legal firms of almost any size.

Info Globe, a subsidiary of the Toronto Globe and Mail, was the first North American vendor to offer a full text news database, with full-text free-text search capabilities. The Canadian content of its product is unique, and the product is competitive. QL Systems has just started to offer a news database called NEWSTEX on behalf of the Canadian Press, an organization owned by over 100 Canadian newspapers.

The news database sub-market is currently quite small because the databases are used by specialists as a business information research tool. Real growth in this sub-market is dependent upon use of such databases by business firms at large, and finally by the individual consumer. Infomart claims that the "Electronic Publishers" will have a major role to play in the creation of this mass consumer market, and that Telidon, operating in the "Gateway" mode, will be the key access tool.

9.5 The Mass Consumer Market

The Canadian mass consumer market is currently at an embryonic stage. Project Grassroots (Appendix 5), operated by the Manitoba Telephone System and Infomart, is the only commercial service currently being offerred to this market. A number of Telidon field trails are underway or planned (Appendix 5), but these are unlikely to be succeeded by significant commercial operations

before late 1983 or 1984. The Source and Compuserve Inc. have a small Canadian clientele, but no information is available regarding the size of their Canadian subscriber base and usage.

Many service suppliers seem to be acting on the assumption that this market will become an important one before the end of the 1980s. They include the telephone and cable television companies, who see themselves as suppliers of Telidon based services to this market, especially to the home. On the content side, the most significant Canadian entrant in this market is Infomart, a subsidiary of Southam and Torstar. Infomart defines its business as "Electronic Publishing", i.e. the use of computer/communications systems to distribute information and transactional services to mass audiences.

Both the Department of Communications and Infomart have been pushing the Telidon system and Telidon-based services as an answer to the problem of penetrating the mass consumer market. Some early estimates of Telidon penetration suggested that 500,000 terminals could be in use by 1985. In the light of the British experience with Prestel, these estimates may be unduly optimistic. The uncertainty lies on the demand side, rather than the supply side. There are no reliable estimates of the size of the mass consumer demand for such services, the types of services demanded, and the price elasticity of the demand.

10. Flows of Public Information Services

(1) With the exception of credit and certain financial information, much of the information contained in these online databases tends to be in the public domain. Trade in packaged databases and retrieval services is therefore more likely than trade in the raw data. One should distinguish, however, between the importation of machine-readable databases by a vendor (like I.P. Sharp, or CISTI) who then builds a public information service around the use of the database, and the direct importation of information services by end users. Importation of databases by the vendor may be inevitable in many instances, given the relative weakness of the Canadian content creation and publishing industries. Direct importation of services by the end user has a greater impact on both the balance of trade and the markets available to Canadian vendors.

(2) U.S. information producers and vendors would, in general, enjoy economies of scale compared to their Canadian counterparts. The vendors also benefit from generally lower computer equipment and telecommunications costs, which form an important part of the total costs of public information services. It is, therefore, easier for a U.S. vendor to serve selected portions of the Canadian marketplace using a computer facility located in the U.S. than vice versa.

(3) The reference database area has a high proportion of imports, both with respect to databases and services. Nine of the 18 databases offered via CISTI's CAN/OLE and CAN/SDI systems are produced outside Canada.

Moreover, retrieval services provided via the U.S. search services, especially DIALOG and SDC, are heavily used by Canadian libraries. It is important to remember, however, that the total size of this sub-market is quite small, perhaps $5-7m. Even if 80% of the services in this area are imported, rather than the 60% assumed in Section 9.2, this would result in a negative trade balance of only $4-5.5m. No information is currently available about the rate of growth of imported services in this area.

(4) Given the existence of CANSIM, Canadian users are unlikely to make major use of U.S. socio-economic databases for information retrieval purposes, although the services of vendor firms such as DRI could be used for economic forecasting and modelling. Due to the activities of I.P. Sharp Associates, which derives two-thirds of its total revenues · (some $24m in 1980/81) from exports to the USA, Europe and the Far East, the balance of trade in this sub-market could very well be positive at this time.

(5) Stockbrokers and financial analysts, on the other hand, must make continuous use of online information regarding the activities of the major American exchanges, provided by vendors like Quotron and Dow Jones New/retrieval, but this type of activity is a condition of doing business and this application has many of the characteristics of a closed user group. No data are available, but it is quite likely that currently imports far outweigh exports of services in this sub-market.

(6)   Credit databases are a major area of potential concern, because of
      privacy and sovereignty as well as economic considerations. We suspect
      that much personal data on individual Canadians, collected by credit
      card companies and financial services, is held in U.S. databases which
      support the North American or even worldwide operations of these
      companies. Examples would include VISA, Mastercard and American
      Express operations, as well as those of finance companies like
      Household Finance. In this area, the need for regulation based on
      privacy and sovereignty considerations may become more important than
      the purely economic issues related to importation of services.

(7)   Canadian users are unlikely to make major use of U.S. legal databases,
      and vice versa. On the other hand, it is quite likely that U.S. news
      databases, like the New York Times Information Bank and the Dow Jones
      News/Retrieval Service, are more used in Canada than Info Globe is used
      abroad. The balance of service flow in the news sub-market is probably
      negative. This sub-market is small now, but could become important in
      the future if full text news databases win mass consumer acceptance,
      and are delivered via Videotex/Telidon services operating in the
      "Gateway" mode.

(8)   While reliable estimates are not available, it is possible to speculate
      that the overall balance of trade between Canada and the U.S. in
      packaged databases and information services lies in the favour of the
      U.S. However, the problem should be viewed in its proper perspective.
      The size of the Canadian market for Public Online Information Retrieval
      Services is currently quite small; some $20 - 45m in 1980. Even if 60%

of the total market was served by imports (and this seems to be a high figure), the negative trade balance would still be only $12-27m. This is only 3-5% of the $530m of EDP services estimated to have been imported by Canadian users in 1980 (DOC Growth Model, Revised Estimates). The negative balance in public information services would be comparable, however, to the estimated positive balance of trade in computer services purchased from the computer service industry. For 1980, the Growth Model assumes service supplier exports of $60m and imports of $30m, for a positive balance of $30m. There is, however, the problem of double-counting the revenues of service suppliers like I.P. Sharp in making the above comparison.

11. Industry Concerns/Views

Industry concerns will be divided into two groups: those voiced by information producers (all the ones interviewed were public agencies), and those voiced by database vendors, both public and private.

11.1 Information Providers/Producers

The two organizations interviewed, Statistics Canada and Environment Canada are both information providers and producers. In one case the data is produced as a part of the agency's primary mandate, while in the other it is an important byproduct.

The primary concern of both agencies is the widest possible dissemination of the information produced. The CANSIM division of Statcan has tackled the problem by housing its CANSIM Main Base at one host service bureau, and making a standard subset of the data, the CANSIM Minibase, available to a number of Canadian and foreign vendors, for a relatively modest annual fee of $20k (Appendix 3). Environment Canada has chosen to house its WATDOC group of databases at a single private Canadian vendor (Appendix 2), but also makes some of the data available through an international database sponsored by the Food and Agriculture Organization, the Intergovernmental Commission of UNESCO and the Ocean Economics and Technology Office of the United Nations Department of Economic and Social Affairs.

Cost recovery and payback considerations can create a major conflict of interest for the public information producer. On the one hand, the producer

may find it highly desirable that data produced with public funds be made accessible to Canadian users via Canadian vendors. But if Canadian market demand is insufficient, in a particular case, to make the database financially viable, the producer may have to subsidize a private vendor by absorbing a portion of his computer and storage costs. A U.S. vendor, operating in a larger market, may not need such financial support and could also provide wider access to the data. The Canadian producer is therefore left with the choice of subsidizing the Canadian vendor, or making Canadian data produced with public funds available to a U.S. vendor for distribution in Canada at no cost to the producer. There are no guidelines to cover this situation when it arises.

11.2 Database Producers/Vendors

All the private database vendors interviewed were opposed to any TBDF policies or restrictions which could affect their ability to offer their products and services outside the Canadian market. This view was expressed most strongly by I.P. Sharp Associates Ltd. and Info Globe. Both these companies have products/services which are competitive in international markets (see Appendix 2,4). They see an integrated North American market for their services, made possible by the integrated Canada/U.S. telecommunications network, combined with important actual or potential markets in Europe and the Far East.

I.P. Sharp Associates, which already does a large volume of business in Europe and the Far East, noted with concern the increasing telecommunications cost of serving these markets with computer facilities located in Canada. Outside Canada and the U.S., the rise of public packet switched networks seems to be accompanied by a trend towards sharply increased tariffs for leased

private lines. Such a development would increase the costs of the private data
networks (packet switched or otherwise), which are offered by vendors like
Sharp, DRI and GEISCO (Appendix 1), to the extent that they may no longer be
competitive with the public data networks. Although the costs of the
telecommunication network are absorbed by the vendor, they must ultimately be
passed on to the user in the form of higher access charges, either accross the
board or differentially for users outside North America. If this trend
develops, a North American vendor may be forced to locate computer facilities in
Europe, in spite of the extra cost of duplicating and controlling databases at
multiple computer sites, to remain competitive in the European market. The
situation is further complicated by access restrictions for U.S. (and Canadian)
database vendors to the Euronet DIANNE network.

All the database vendors interviewed insisted that they must be free to
import databases produced outside Canada, and offer them to both Canadian and
non-Canadian users of their services. This is seen as an essential requirement
in serving both the Canadian and foreign markets, because a large proportion of
the data demanded by the users is produced outside Canada.

In addition to these global TBDF-related concerns, others were
expressed which were specific to particular vendors. One private vendor claimed
that competition from publicly funded agencies (federal or provincial), often
operating on a less than full cost recovery basis, served to reduce the market
effectively available for private vendors. This has not been a concern as yet
in the source-numeric sub-market. In this context, it should be noted that the
Office of Management and Budget of the U.S. federal government has (in a
memorandum dated September 11, 1981) requested all government departments and

agencies to determine whether federal information centres or activities duplicate private endeavours or pay their own way.

Concern was also expressed by database vendors that the federal government is involved in database activities in-house which the private sector is able to provide. It was felt that at the least, the federal government could apply its "make or buy" policy more rigorously in this area.

Another vendor expressed the view that the Department of Communications may be overemphasizing Telidon as an information delivery technology, to the detriment of other established and developing technologies for distributing public information services. It was pointed out that the Telidon data structure, built around the concept of the "page" and the tree-search, was neither necessary, or even particularly suitable, when it came to dealing with conventional source numeric, source full text or reference databases. The transactional and gateway capabilities of Videotex systems were, however, seen as two very important factors in wider use and spread of this type of public online information service.

12. <u>User Viewpoints</u>

The interests of the end user are not necessarily the same as those of the Canadian supplier (information producer or online vendor), and must be kept in mind when considering trade in, or importation of services. The user requires appropriate information, at the least cost, to improve his efficiency and productivity; if this is not available from a Canadian supplier, he will go to a foreign one. In many cases the gains to the user, in terms of improved efficiency and productivity, may far outweigh the loss to Canadian suppliers, represented by the importation of the information service.

Time and resource constraints of this phase of work of the Task Force did not, unfortunately, permit for a survey of user's views on access to information in the context of TBDF. The only users that could be readily contacted during this phase of the project were a group of reference librarians in key federal departments and organizations (see Appendix 6). All these libraries operate some form of a computer based reference service for their users, accessing both Canadian and U.S. database vendors. The librarians were unanimous in pointing out that access to the U.S. vendors (Dialog, SDC, BRS and NLM) is an essential requirement for the service that they provide to their clients (primarily researchers). Several also expressed the view that it should be left to the information specialists' discretion to decide how a user request could be best serviced, instead of imposing access priorities.

In the context of other forums and studies, users (especially the scientific research community) are quite adamant that access to data or information should not be restricted or hindered in any way.

13. <u>Economic Issues</u>

[Draft needs major revisions as a result of changes to Sections 5, 10, 11 and 12].

14. Culture Sovereignty and Transborder Data Flows

Public online retrieval services are more than just value-enhancing data processing services, these services with their data bases have a cultural dimension as well. While general cultural aspects and sovereignty in the context of TBDF are discussed in another report, there are a number of specific questions which should be addressed here.

14.1 Cultural Exports and Imports Act

Current Canadian policy on the import and export of culture (artifacts, documents, etc.) is set forward in the Cultural Property Exports and Imports Act. However, this act applies to hard-copy records only and not to those in machine-readable form. Further, the act applies only to records which are over 50 years old. Consequently, the act has no relevance to the question of ensuring an electronic archival or cultural heritage. There have been occasions where Canadian data bases or data base activities, i.e. those with primarily Canadian content or Canadian focus, have have become submerged or swallowed up into larger databases existing elsewhere.

The question here is two-fold:

a) as artifact and information merge into one electronic digitized form, with an age factor being a rather meaningless concept, should or can principles embodied in the Cultural Property Exports and Imports Act be applied to the machine-readable or electronic data?

b) are data bases or data banks to be considered cultural entities? Are there certain types of data bases which are considered to be "culture free" i.e. scientific or technical and others which are heavily culturally weighted, i.e. social sciences and source data bases? Is it possible or will it be a useful exercise to classify or grade data banks as cultural entities of note?

14.2 Cultural Industry

The Canadian government has always been actively involved in the maintenance of a cultural industry be it in the form of support and stimulation through various grants and research programmes, through regulations put forward by the CRTC or through various tax measures such as those relating to film making (Canadian Film Development Corporation) and Canadian advertising in foreign countries (Bill C-58). No discussions have taken yet place on the desirability or applicability of these mechanisms to data bases and data files as a cultural industry. Certain videotex (Telidon) data bases could be viewed as art forms as well as information providers. This raises the question of government involvement/support in the creation of electronic art forms. This raises the question of whether databases and data files could or should be considered as another (new) sector of the Canadian cultural industry.

A number of current government programs related to the cultural industry should be noted here as they may provide useful analogies.

a) The ART Bank of the Canada Council through its purchase of works of Canadian artists seeks to ensure the viability, development and continuance of an indigenous artist community/industry.

b) The Scholarly Publication Program of the Social Sciences and Humanities Research Council subsidizes the cost of publishing Canadian works of scholarly merit.

c) The Department of Communications has sponsored the creation of a Telidon-based information retrieval service in the form of a data base of Canadian records, hoping thereby to stimulate the Canadian recording industry.

In the creation of both data bases and data files, as much as 90% of the total cost lies in creating the machine readable data in the first place. Putting the data base up (plugging into a distribution network) or cleaning and reformatting data files (magnetic tapes) for general distribution represents only 10% of the total costs. Quite often those who create the data base or data files do not budget for this 10% since the prime purpose for the creation of machine readable data is to support a certain program activity or research project, e.g. publication of federally funded research on the disabled-as part of Canada's contribution to the 1981 International Year of the Disabled. Even those who feel that their database or data file is of general interest and can make a significant contribution to our knowledge of Canada are prevented (economically) from distributing this knowledge electronically, either because the distribution of the data is not seen as a commercially viable operation or

those who supported the creation of the data do not have either the financial resources or mandate (or desire) to fund such a distribution.

The question to be raised in this context are, "Is the support of 'electronic' publishing either through on-line retrieval services (data bases) or preparation and distribution of data files and their related documentation (magnetic tapes) an activity the government wishes to give consideration to?" and "Is the timing right for such a consideration?"

14.3 Inventory of Data Bases and Data Files

Various mechanisms are currently in place which assist in the identification or inventory of various components of the Canadian cultural heritage. For published works, the requirement of a legal deposit in national and/or provincial libraries ensures that this hard-copy Canadiana in print form is identified, conserved and made accessible (in varying degrees) to all Canadians. This system also promotes a positive form of TBDF from the Canadian point of view, allowing those in foreign countries to ascertain quite readily what Canada has to offer in books, periodicals and near-print literature.

The Telidon project of the Department of Communications for Canadian records fulfills a similar function for the records industry. Other efforts of note are the project of National Museums for a comprehensive Canada-wide computerized inventory of artifacts and a similar project for films by the National Film Board.

There are no such concerted or comprehensive efforts currently being
made with respect to what might be termed our "electronic cultural heritage".
Nevertheless, two activities currently exist which could provide a basis for the
identification and inventoring of data bases and data files. These are the COIN
Directory of Computerized Information in Canada, sponsored by the University of
Alberta and the Automated Inventory of Canadian Machine Readable Data Files,
developed by the Machine-Readable Archives Division of the Public Archives of
Canada.

The COIN Directory is now in its third edition and has become the most
authoritative and comprehensive guide to public data bases currently accessible
in Canada; it operates as an on-line data base at the University of Alberta
running under SPIRES software. While not addressing the question of data
ownership directly the COIN Directory does identify the source, i.e. the author
or organization that prepared the data base (See Appendix 2 for a compilation of
data on databases and database vendors taken from the COIN Directory and other
sources). The COIN Directory is considered the most comprehensive and extensive
inventory to date of data bases available to Canadians and being used by them.

The idea of a data base of data bases or, more precisely, a data base
of Canadian data bases has been advocated for some time both within government
circles (e.g. the National Library, the Canada Institute for Scientific and
Technical Information (CISTI), DOC's Telidon Project) and in the social science/
research community as well. However, to date various initiatives have consisted
of one-shot projects or continuing efforts in a single subject area only. While
the COIN Directory lacks information which would be especially useful for

analysis from the TBDF point of view, modifications to its program appears to be
the most effective and cost-beneficial approach to the establishment of an
on-going, online database on databases accessible to or being used by Canadians
with as a sub-set a database of Canadian databases. Criteria for defining the
latter need to be refined and will require further consultation with industry
and affected institutions and individuals. The establishment of an optimum and
mutually beneficial working relationship between the public and the private
sector also warrants some attention.

The Automated Inventory of Canadian Machine Readable Data Files is a
one year old project of the Machine Readable Archives Division of the Public
Archives of Canada. It is an inventory of machine readable data files (magnetic
tape) known to exist in Canada. Comprising information on approximately 5,000
data files, the inventory has information on data files existing within both the
federal government sector (about 3,500) and those in the private and
not-for-profit or academic sector (about 1,500). Finally, 'it should be noted
that the Federal Database Group of Statistics Canada maintains a computerized
inventory of all federal information banks (hard copy and machine readable)
which have to be registered according to provisions of the Canadian Human Rights
Act and Treasury Board Canada regulations.

Proposed Access to Information and Privacy legislation (Bill C-43) will
require government agencies to identify their information holdings. Pilot
projects (sponsored by the Treasury Board, Statistics Canada and the Public
Archives) are currently underway which address specifically various aspects of
the application of Bill C-43 to machine readable data. Consequently, there
exists a real possibility that an inventory of Canadian data files at least

those pertaining to federal government entities will become an on-going activity and reality. With respect to including non-federal government data files in the inventory, further discussions with the Public Archives and various data archives (public and private sector) across Canada would be in order. (The Public Archives has had, for some time, the creation of a Union Catalogue of. Machine Readable Data Files as a long-term objective. Perhaps similar work by the Museums Canada and the National Film Board can provide some guidance in this respect).

Only with an up-to-date database on Canadian data bases and data bases accessible to Canadians and an Union Catalogue of Machine Readable Data Files will it be possible to analyse and identify supply trends in machine readable data as a cultural industry and cultural heritage. Such information is a prerequisite for being able to gauge both the qualitative and quantitative aspects of TBDF as they relate to data bases and data files so that the government may be able to identify specific problems and take remedial or stimulative actions when appropriate.

14.4 Cultural Dependency and Transborder Data Flows

Very little hard data is readily available which would make it possible to discern the cultural impacts of transborder data flows involving public on-line information services. Criteria are lacking for evaluating what data bases are relatively culture free and which have a high culture content. (Some would even question the validity of such an undertaking).

Nevertheless, survey data that is available does indicate a heavy dependence on foreign data bases (mainly U.S.) by Canadians. A survey by Gilles Deschatelets (1979) on online search services centres, indicated a clear domination of U.S. data bases and U.S. vendors. His findings showed that of the 380 online service centres surveyed availability of online systems at these service centers was:

(Top Five Systems)

ORBIT	- 24.5%	(U.S.)
CAN/OLE	- 17.7%	(CDN.)
DIALOG	- 17.4%	(U.S.)
QL SYSTEMS	- 14.0%	(CDN.)
MEDLINE	- 6.3%	(U.S.)

These figures indicate that ORBIT, a U.S.-based service marketed in Canada by Informart Ltd., has the highest market penetration being available at 24.5% of the online service centres surveyed.

According to Deschatelets in terms online systems usage, the top four systems were:

ORBIT	- 35.9%
DIALOG	- 27.1%
CAN/OLE	- 22.9%
QL SYSTEMS	- 14.1%

A ranking of most frequently queried databases by Deschatelets shows that the top ten data bases were all non-Canadian. They are, in ranked order:

COMPENDEX

CHEMLON

NTIS

INFORM

ERIC

BIOSIS

PSYCH. ABST.

MEDLINE

INSPEC

MANAGEMT

Another study by Stephen Lawton in (late 1979) on online retrieval in post-secondary institutions also noted that the top ten most searched databases were non-Canadian. Lawton gathered some preliminary data on payments made to online systems and while the figures are tentative, they do show that out of each $100 spent on public online information retrieval services in the post-secondary institutions,

DIALOG (U.S.) received — $47;

ORBIT (U.S.) received — $19;

BRS (U.S.) received — $13;

MEDLINE (U.S.) received — $11;

CAN/OLE (CDN.) received — $ 5; and

QL SYSTEMS (CDN.) received — $ 4.7;

This would lead to a tentative conclusion that, at least in the reference database sub-market of online information retrieval services, Canadian dependence on TBDF for delivery of this service could be as high as 90%.

A sampling of a select number of major federal libraries on their use of online services yields an additional insight into usage and flow patterns. One is a ranking of the top ten services according to the number of connect hours, the other a ranking according to monies spent on using the services.

[NOTE: Tables are not ready since returns are still coming in.]

Preliminary analysis of the databases available and accessed by Canadians (Appendix B) indicate that sources of unique reference and bibliographic bases are

Canada	– 35%	
U.S.	– 56%	(out of 288)
England	– 5.7%	
Netherlands, France and Switzerland	– just under 1%	

And similarily for source-numeric data bases, the origins are,

Canada	– 66%*	
U.S.	– 28%	(out of 68)
England	– 4½%	
Austria	– 1½%	

(* I.P Sharp is the major factor here).

14.5 Cultural Identity and Sovereignty

It has been estimated that about 3% of the world's research is done in Canada (this is not that disproportionate considering that Canada, with its heavy dependence on trade, accounts for about 5% of the total world trade). Consequently, there is a built-in (economic) motivation to add knowledge of this 3% to the 97% for the purpose of data base creation and global dissemination/ distribution purposes. This is not a unique situation and could be considered the norm looking at other sectors of Canadian culture.

In the public on-line retrieval sector, one could work towards proportional equality and reciprocal sharing. Several possibilities exist.

a. World Product Mandating or Specialized Missions

In order to ensure a larger degree of reciprocity in international trade, to encourage a larger degree of independent action and flexibility as well as a more equitable distribution of research and development work, multinationals have in recent years adopted the concept of world product mandates or specialized missions for their subsidiaries in various countries. Multinationals and their national subsidiaries and the host countries view world product mandating as a very practical and realistic approach to ameliorating a number of conflicts that have arisen in the relation of the multinational corporation to the nation state.

A number of data bases are created and maintained in the context of cooperative international arrangements. The motivation for such activity is basically two-fold, namely:

- the need for data base (reference or source) is identified and recognized simultaneously by many nations usually in the context of some on-going joint international activity; and

- individual participating entities (national and/or private) decide that the practical needs of each will be served best by pooling effort and resources in order to build a data base which is as comprehensive (authoritative) as possible.

Examples are data bases such as the FAO agriculture data base in Rome the IAEA atomic energy data base in Vienna and the ILO labor data base in Geneva. Participating countries forward their data to those processing centres where the master data base is created. One can either access the data base on-line or receive copies of the up-dated data base for distribution on national or other networks closer to home.

In this context, the question is raised whether the federal government and the private sector would find it desirable to identify some possible subject areas for which one may wish to obtain a world product mandate or specialized mission. Such specialized missions in database activities could be linked to vital information needs of Canadian government industry and the research community. Example, which might be considered is that of arctic/northern research pulp and paper, or telecommunications. However, it should be noted

that the idea is presented here as a concept only and further research needs to be done on the practicality and other ramifications of adopting such a strategy should there be consent that the concept has merit.

b. Canadian Identity in Data Bases

Canadian identity, in terms of Canadian contributions to the content of data bases can be maintained in two ways. Either one physically creates Canadian data bases or one ensures the existence of a logical "Canadian content" data base within a large data base. In this section, we address the latter point.

For the most part reference databases focus on a specific subject, e.g. education, toxology, chemical engineering, etc. The purpose of such data bases is to capture as much data on a world-wide basis on that particular subject, thereby ensuring a "comprehensive and authoritative and thus commercially viable database. Queries of scientific and technical databases are usually of an acultural nature, e.g. "What literature is available on the use of barium crystals for low-energy lasers?" But if one queries an international data base for non-geographic specific information such as, "What have Canadian doctors written on medical ethnics and artifical insemination?" or "What are examples of methodology used by Canadian social scientists to study North-South relations?", it is difficult to obtain an answer. For most scientists the nationality of the author is of little or no concern. Social scientists are not so sure. Given the fact that Canada's bargaining power internationally vis-a-vis public online information retrieval services is of the order of 3-5%, it still may be possible to ensure that the Canadian identity in international data base is maintained by

adding a field tag or "flag" to each record identifying the author as Canadian. In this sense, a Canadian identity would be maintained logically in such data bases.

Once again this is a concept which needs further exploration in terms of both practicality, modes of implementation and costs.

c. Promotion of Canadian Data Bases

The federal government has actively promoted creation and maintenance of a Canadian cultural identity through policy and regulatory actions and stimulative measures. Due to the fact, that very little is known as yet on the demand or consumption side of the data bases and data files, the discussion here will limit itself to supply-side possibilities only.

A substantial part of research in Canada is supported by public funds (federal government) and in many cases involves the creation of machine readable data. However, in terms of reporting on the products of such research by where, the researcher, contractor or the contracting federal agency, the situation is out of control. The mechanisms to do so are weak or non-existant, i.e. those receiving federal government funds or those dispensing the same for activities which generate data bases or data files do not (are not required) to report on these products to any information "clearing house(s)". Even where federally-funded (research) contracts state that "all material created as a result belong to the Crown", the reporting, retrieval and eventual dissemination/distribution mechanisms are inadequate.

Of importance to the TBDF issue is the fact that this type of data forms a very unique source for unique Canadian content to reference and source data bases. Consequently those creating online Canadian data bases (in both the public and private sectors) are deprived of raw domestic input which, if captured or incorporated, would enhance both the economic and cultural value of such data bases.

The situation is much the same for data files created either as source data or inventories/bibliographies on a particular subject. In both cases, the downstream economic possibilities for on-line information services and the possibility of systematic augementation of Canada's electronic cultural stock are lost.

This raises the question of whether a more rigorous and systematic application of current government policies in ensuring that the existence of the research results, i.e. those in machine readable data, resulting from the expenditure of federal funds are made known throughout Canada and the world would faciliate the use of the same as raw data input for the Canadian online services industry thereby promoting some outflows some of which could be profitable or at least revenue-generating.

As in other cultural areas, support is often given for those activities with a high cultural content which cannot survive if left totally dependent on the economics of the marketplace. The same holds true for Canadian data bases. As stated above as much as 90% of the cost is in creating the data base while as little as 10% may be the cost of putting the data base up as an on-line service thereby ensuring coast-to-coast distribution. Drawing an analogy to both the

Aid-to-Scholarly Publications Program of the SSHRCC and industry stimulation programs, (such as those for Telidon, those of DREE and government budgetary tax policies), a number of options or courses of action with respect to database and data file creation as a cultural industry present themselves:

(1) Measures involving Taxation Policy:

Tax Incentives currently exist for individual taxpayers to engage in the production of Canadian movies. There is also the long-standing industry request for different forms of tax relief related to R&D expenditures.

(2) Stimulative Programs

The Department of Communications is an effort to stimulate the development of a Telidon-based industry introduced a one-for-one program. The Department of Regional Economic Expansion provides assistance to establishment of new industry in economically disadvantaged areas.

(3) Direct or Indirect Assistance Programs

The Aid-to-Scholarly Publications Program of the SSHRCC provides financial assistance through subsidizing that part of the publication cost of works of scholarly merit which Canadian publishers estimate would not be covered by market sales. The Department of Communications

Telidon database on Canadian records is an example of indirect economic-cultural stimulation of the Canadian record industry.

In this context, a number of possibilities were noted such as:

(1) The establishment of a fund or program to provide seed money to generate Canadian database activities which either themselves have a good chance of becoming a commercial success or which would stimulate other economic-cultural activities.

(2) The establishment of a fund or program which would assist in the cleaning or reformatting of machine readable data files of a cultural (and academic/research) value thereby promoting their dissemination and distribution.

(3) The establishment of a fund or program which would ensure that Canadian databases of cultural importance are ensured widespread dissemination and distribution through support of that particular expense of the whole activity (i.e. the "10%").

Alternatives such as these were noted as possible measures which would promote the economic vitality and cultural uniqueness of the Canadian on-line information retrieval industry. The more unique Canadian products generated, the greater the possibility for exports of data flows since examples of successful new ventures in on-line information services either the unique content or "the most authoritative content" forms the key to successful operations.

14.6 Access Protocols and Standards

Questions have been raised as to the possible effect of standard access protocols and dialogue procedures (i.e. U.S., English) for francophone Canadians apart, from a French-language database content itself and how this may tie into official languages policy and language rights under the new constitutional charter. These questions while valid fall outside the domain of the TBDF Task Force. Nevertheless, a number of observations can be made.

a. The systems operated by QL Ltd., Can/Ole, IDRC and _____ allow for access procedures and dialogue in either English or French. The problem is therefore less acute for Canadian on-line information retrieval services than it is for those which are imported and where English is the predominant (if not universal) access language.

b. In the early days of the industry, the use of free text vocabularly and full text data entry was more costly than the use of key words and abstracts. Users were therefore dependent on the language deployed in the keywords and abstracts in defining search strategies. The introduction of full text data bases and free text vocabulary with bilingual or multilingual thesaur have brought the language question back to that of the reference title or document itself. In that sense the computer has become language blind as it works with unique (whole or truncated) character strings. However, a thorough analysis of the interplay of a very strong universal tendency towards the use of English (U.S.) versus an increasing

trend towards the use of national languages in terms of access and operating system institutions and the resulting cultural impacts fall outside the scope of the work of this project.

Nevertheless, some recent experience with Telidon (Ref. D11) in this area noted linguistic problems related to:

Error Messages: (Even when the user's access protocols and dialogue are in French some of the error messages or the more detailed trouble-shooting manuals are in English only).

Character Sets and Diacritics: (Databases often get around the problem of diacritics by using upper case only but even then problems do present themselves as there is a difference between UN HOMME TUE and UN HOMME TUÉ. Further on the data side when using lower case, the entry of French characters with diacritics can require simultaneous combination of keys which is more time-consuming and error-prone (and therefore more costly) than being able to use a "hard wired" bilingual keyboard.

Translation: (The possibilities of translation range from word-to-word to literacy translations. For source databases such as full text legal databases this is less of a problem than for reference or source databases which still rely on keywords (which often do not lend themselves to an exact translation).

c. The harmonization of standards and protocols in terms of a set of technical specifications and interfaces would allow for the interconnection of on-line data networks. Presently, communication protocols in the computer industry are proprietory while those in the communication industry are not.

Lack of national standards would hinder the development of national networks while differing standards between nations would

- affect (hinder) transborder data flows;
- add significant costs to the user;
- foster private/closed networks; and
- hinder the development of an "open public inter-connect network".

While some of the data base vendors of source data operate private networks (e.g. I.P. Sharp), most other data base vendors especially those selling reference data bases rely on public networks (i.e. the telephone system with DATAPAC, Tymnet, Telenet) for their distribution. It might be noted here that the National Library has recently started a major networking project to test new systems for exchange of bibliographic information taking the approach of the open systems interconnect model, i.e. by participating in Bell Canada's iNet project.

15. Legal and Institutional Issues

15.1 Legal

A decade ago much of the initial work in establishing the on-line reference data bases was carried out by librarians. They identified possible applications for automated support services in providing reference data bases. Consequently, many of the reference data bases consist of compilations of information that is considered as being in the public domain, i.e. titles of articles, books, reports, etc. In a number of cases, a particular library, organization or institution undertook as a project, the preparation of abstracts or precis of these publications. The area of coverage of the reference data bases usually coincided with the thematic mandate of the library or of the organization which it served.

Concurrently, certain publishers and nascent information companies started developing on-line data bases as a commercial product using both public and proprietary information. The themes or areas of concentration chosen were those which either promised the best commercial viability or those for whose early development someone else (usually a government agency) had provided the initial start-up costs. The net results of the early history of the development of on-line data bases in both the public and private sectors is a melange of property rights with respect to ownership of information in these data banks.

At the same time, no clear dominant pattern has yet emerged in the relationship between information creator, processor, vendor, distributor, etc. In some cases, the vendor buys a data base with or without exclusive

distribution rights. In other cases, a data base creator pays a processor to build the data base and a vendor to distribute it with or without royalty charges. In some cases, the processor, vendor and distributor are the same. Many different combinations of "data ownership" with their ensuring legal complications are possible.

In addition, data bases originating out of the library world (which traditionally does not charge for information) often became part of larger cooperative efforts to establish a larger (more authoritative) data base on a specific subject, the arrrangement being covered by no more than a handshake or exchange of letters. As a matter of fact, the concept of "data ownership" has only recently become a real issue now that the industry has passed its initial development stage.

As the on-line information industry matures, the need for a sound legal framework is becoming more pressing. There is unanimous consent in the industry that the lack of a clear legal framework for data ownership rights is a serious impediment to the development of the industry. Those in the public sector discover that their "hand shake" arrangement in a cooperative venture does not prevent such developments as take-overs now that the data base they nurtured has become a (commercial) success.

The situation is less complicated for source data bases where one-to-one (data base creator/producer to vendor/distributor) relations seem to be the norm. In a number of instances, a single organization performs all these functions. Nevertheless, here also concern has been expressed about the lack of a clear legal framework.

For both the reference and source data bases in the public and private sectors, a revised copyright act and amendments to the criminal code are looked upon as the cornerstones of a legal framework for the protection of property right with respect to data or information. Industry hopes that amendments to the copyright act will address the question of "electronic licensing", i.e. assist in ensuring that just compensation is given for use (or repeated use) and prevent unauthorized "copying" of the information transmitted.

Those interviewed also stated that amendements to the criminal code should provide legal redress against unauthorized use of computer-communication systems in terms of both access to the data and tampering with the data (addition, deletion or modification).

In TBDF context, the resolution of the question of data ownership rights for electronic or digitized data in terms of domestic law should be carried out with as sub-objective the harmonization with laws addressing similar questions in other countries.

Representitives of the public on-line information services industry hold strong negative opinions on the question of the use legal instruments such as regulations in relation to TBDF. For the foreseeable future, the possible benefits of regulations will probably not equal a concomitant loss of flexibility in operations and development of new services. Most on-line information services therefore do not favour (or are against) any TBDF related regulation which could impact adversely on this flexibility. It should be noted that those operating the public sector or academic in library/information services, while stressing the need for continued availability of access to data

anywhere in the world, do see a need for the use of legal instruments as a means for protecting and ensuring Canadian content/culture or a data processing software capability in Canada. (For further details see the other chapters in Section D).

E. Conclusions

(Currently, Note 1. The conclusions will be put in regular text form pending review of comments by industry and other Task Force participants).

Appendices

As stated in Note 1 a number of appendices will be forwarded as soon as ready.

References

A. General

1. Telecommunications and Canada
       Report of the Consultative Committee on the Implications of
       Telecommunications for Canadian Sovereignty, March 1979.

2. Darrow, J.W. and Belilove, J.R.
       Keeping Informed: The Growth of Databank Sharing
       Harvard Business Review, Nov.-Dec. 1978

3. Kiechel, W.
       Everything You Always Wanted to Know May Soon Be On-Line
       Fortune, Vol. 10 (3), May 1980

4. Bourne, C.P.
       Online Systems: History, Technology and Economics
       Journal of the American Society for Information Science,
       Vol. 31 (3), May 1980

5. Wanger, J. and Landau, R.N.
       Non-bibliographic On-Line Database Services
       Journal of the American Society for Information Science,
       Vol. 31 (3), May 1980

6. Lawton, S.B.
       Diffusion of Automation in Post-Secondary Institutions
       Canadian Library Journal, April 1981

7. The Library and Information Manager's Guide to On-Line Services
       Ryan E. Hoover (Ed.)
       Knowledge Industry Publications, Inc., 1980

8. The Electronic Library: Bibliographic Data Bases, 1978-79
       R. Christian
       Knowledge Industry Publications, Inc., 1978

9. Anderla, G.
       Information in 1985: A Forecasting Study of Information Needs and
       Resources
       OECD Report, Paris, 1973

10. Clayton, A.
       Factors affecting future online services
       Online Review, Vol. 5, No. 4, August 1981

11. Lancaster, F.W. and Smith, L.C.
       Online Systems in the Communications Process: Projections
       Journal of the American Society for Information Science, Vol. 31(3),
       May 1980

12. Micros: A Pervasive Force
    M. Orme
    John Wiley and Sons, 1979

13. Policy Framework for the Economic Development of the Communications/Information Sector
    Report prepared for the Department of Communications by Price Waterhouse Associates, November 1981

14. Deschatlets, G.
    A survey of Online Search Service Centres in Canada
    Laval University, March 1980

15. Lorenzi, J.H. and Toledano, J.
    The International Online Information Retrieval Market: An Economic Analysis
    OECD Report, DSTI/ICCP/81.28, November 1981

16. Computing Trends: A Review of Computer/Communications Technology
    Report prepared for the TBDF Task Force, December 1981

17. The Growth of Computer Communications in Canada
    Computer/Communications Secretariat, DOC, March 1978
    Revised Estimates and Projections to 1990
    Report prepared for the TBDF Task Force, December 1981

B. Producers and Vendors of Public Information Services

1. Encyclopedia of Information Systems and Services
    A.T. Kruzas and J. Schmittroth, Editors
    Gale Research Co., Detroit, Mich., USA
    4th Edition, 1981

2. Computer-readable Data Bases: A Directory and Data Sourcebook
    M.E. Williams et. al., Editors
    American Society for Information Science, Washington, D.C., USA
    October 1979

3. Directory of On-line Databases
    R.N. Landau et. al., Editors
    Cuadra Associates Inc., Santa Monica, Calif., USA
    Vol. 2, No. 1, Fall 1980, Vol. 2, No. 3, Spring 1981, Vol. 3, No. 1.,
    Fall 1981

4. COIN: A Directory of Computerized Information in Canada
    S. West, Editor
    Alberta Information Retrieval Association, Edmonton, Alta.
    1st Edition, March 1980, 2nd Edition, June 1981

5. The Espial Data Base Directory
    H.C. Campbell, Espial Productions, Oct. 1979

6. CAN/SDI: Canadian Service for the Selective Dissemination of Information
  Canada Institute for Scientific and Technical Information (CISTI)
  National Research Council of Canada

7. CAN/OLE: Canadian Online Enquiry System
  Canada Institute for Scientific and Technical Information (CISTI)
  National Research Council of Canada

8. CISTI Annual Reports: 1977/78 - 1980/81
  National Research Council of Canada

9. QL/SEARCH: The Information Retrieval Service
  QL Systems Limited, March 1981

10. CANSIM Services in Canada
  Statistics Canada

11. Public Data Bases
  I.P. Sharp Associates Ltd., January 1981

12. Public Data Bases Catalogue
  I.P. Sharp Associates Ltd., December 1981

13. Public Databases
  EDP In-Depth Reports, Evans Research Corporation
  Vol. 11, No.2, October 1981

14. Information and Database Publishing Report
  Vol. 2. Nos 1-18. (Nov. 1981)
  Knowledge Industry Publications Inc.

15. Pricing Online Database Use
  LINK Research Report, ORR Vol. 1, No. 4, December 1980

16. Finding New Database Opportunities
  LINK Research Memorandum, ERM Vol. 2, No. 9, May 1981

17. EUSIDIC Data Base Guide, 1981.

C. <u>U.S. Industry Studies</u>

1. On-Line Data Base Services Markets

  International Resources Development (IRD) Inc., March 1981

2. On-Line Data Base Services
  Creative Strategies International (CSI), January 1981

3. Strategies in the On-Line Data Base Marketplace
  Link Resources Multi-Client Study, 1979

4. On-Line Data Base Markets
  Link Resources Report, April 1981

5. International Market Opportunities for Online Database Services
   Input Ltd., Sept. 1980

D. Videotex/Telidon

1. Videotex in Canada
   J.C. Madden, Department of Communications, 1979

2. Gutenberg Two: The New Electronics and Social Change
   Ed. Godfrey, D. and Parkhill, D.
   Procip Press, Toronto, 1979

3. A Study to Forecast the Demand for Telidon Services over the Next Ten Years
   A Report prepared by R.W. Hough and Associates Ltd., for the Department
   of Communications, December 1980.

4. Telidon: The Information Providers
   M. Kurchak, Morgan Communications Ltd.
   Report prepared for the Department of Communications, March 1981

5. Data Banks in Canada: Type, Ownership, Control and Access
   Simon Fraser University
   Report prepared for the Department of Communications, Nov. 1981

6. Telidon Reports No. 1-7
   Department of Communications, Feb. 1980 - Sept. 1981

7. International Videotex Teletext News
   Nos 1 - 12 (May 1981)
   Arlen Communications Inc., Washington, D.C., U.S.A.

8. MICROS: The Canadian Micro-systems Journal
   Vol. 1, No. 3, November - December 1980
   Vol. 2, No. 4, May 1981

9. Videotex 81
   Proceedings of the Videotex Conference, Toronto, May 20-22, 1981
   Published by Infomart and Online Conferences Ltd.

10. Parkhill, D.F.
    Developments on the Canadian Scene
    Paper presented at Viewdata 81 International Conference and Exhibition,
    London, October 6-8, 1981

11. Fortin, J.Y.
    TÉLIDON et la francophonie
    Exposé présenté à la 9e conférence annuelle de l'Association canadienne
    des Sciences de l'information, 27-30 mai 1981.