

A FEASIBILITY STUDY TO ESTABLISH
A DATA BASE
for the
SPACE STATION
USER DEVELOPMENT PROGRAM

DSS CONTRACT NO.
02SW.31098-6-6014

Philip A. Lapp Limited
Suite 302
14A Hazelton Ave.
Toronto
M5R 2E2
March 23, 1987

Table of Contents

EXECUTIVE SUMMARY	i
1. INTRODUCTION	1
1.1 Background	1
1.1.1 A Canadian Data Base	1
1.1.2 The NASA Aerospace Database	2
2. METHODOLOGY	6
3. FINDINGS	8
3.1 General	8
3.2 INIS and Sport Database	9
3.2.1 INIS	9
3.2.2 Sport Database	10
3.3 The Surveyed Data Bases	12
3.3.1 CODOC	12
3.3.2 CISTI Data Bases	14
3.3.2.1 CISTI Catalogue Online (OON)	14
3.3.2.2 IEC	15
3.3.2.3 NRC/PUBS	16
3.3.2.4 Translations Index	17
3.3.3 MICROLOG, Canadian Business and Current Affairs, Directory of Associations in Canada	18
3.3.4 GEOSCAN	20
3.3.5 Canadian Transportation Documentation System (OOT)	21
3.3.6 Remote Sensing On-Line Retrieval System (RESORS)	22
3.3.7 Department of National Defence (DSIS)	24
3.4 Sourcing List - Department of Supply and Services	25
3.5 The Israel - US Agreement	25
3.6 Sample Searches	26
3.7 Summary	27
4. CONCLUSIONS	31
5. RECOMMENDATIONS	36
APPENDIX 1	38
APPENDIX 2	44

EXECUTIVE SUMMARY

Recent Canadian initiatives in relation to space activities, particularly the decision to establish a Canadian Space Agency, have brought to the fore the question of availability of information concerning what has already been learned, both in Canada itself and elsewhere in the world.

Canada has been a leader in space investigation for many years, until recently concentrating on communications technology, space science and related systems engineering. Access to microgravity for extended periods of time can now be contemplated, and much has already been learned through use of the Shuttle and by the Russians through their own space program.

While scientific results have by and large been made available through the normal international publishing channels, there remains a very significant body of knowledge contained in the non-conventional literature - unpublished reports, articles in non-scientific publications, industrial reports etc. - that is not so easily obtainable, and yet remains in the open literature.

NASA has initiated a program to collect all space-related information into one comprehensive database, the NASA Aerospace Database. This has two components, the International Aerospace Abstracts (IAA) and the Scientific and Technical Aerospace Reports (STAR). The former deals with the scientific literature and the latter with the non-conventional literature. The scientific literature finds its way into the public domain through other channels, as mentioned above, but the non-conventional literature must be ferreted out through special networks. For STAR, much comes from the US itself, but NASA also enters into agreements with other countries for supply of their information in return for access to the NASA Aerospace Database.

ESA and Israel have signed agreements already, with others to follow. It is one option for Canada to pursue.

The present study was commissioned to assess the feasibility of establishing a Canadian Space Users Data Base, for use by Canadians in the first instance, but also to be used to enter into an exchange agreement with NASA, if such a course were to be deemed desirable.

Many data bases exist in Canada at present, and a considerable amount of non-conventional information is available through on-line retrieval services. A review was carried out of the contents and sources of the data bases, and ten were selected for more detailed investigation, either because the content seemed to correspond to the requirements of the Space Users Data Base, or because of particular experience the managers had acquired through participation in international data bases.

Managers of the selected data bases were interviewed to verify contents, identify sources, ascertain procedures for data entry, obtain estimates of costs, and learn what was involved in joining international information systems. In particular, they were asked if they would be willing to contribute to the new data base and to assess the impact on their operations if they were to adjust procedures to conform to NASA standards for data entry.

Interviews revealed that a great deal of the Canadian non-conventional literature is being collected, largely from federal government departments, but that the procedures for entering the information vary widely. The purposes for which the data bases were established range from titles with financial information through process for cataloguing to full abstracting and indexing. In some cases, particularly the commercial operation we visited, the procedures are complete - abstracts, indexes, documents held

on fiche. In others only rudimentary information is recorded.

Most of those interviewed were willing to contribute documents, but none was at all enthusiastic about changing the present procedures to conform to NASA standards. Cost estimates vary from about \$15 per entry to about \$50 per entry, the latter representing a realistic figure for the proposed data base.

It was learned that the STAR abstracts are available commercially, DSIS has a subscription, but the process of extracting information from hard copy is cumbersome and time-consuming. For those who want access to the NASA system, on-line service appears to be the preferred route.

The Department of Energy, Mines and Resources has recently released an RFP through DSS to establish a Canadian Energy Database that would serve much the same purposes for the energy community that would the space data base for the space community. Access to the DOE (US) energy information is also restricted to those who contribute their own data.

It should also be remarked that the NASA information has a component that is classified, and that component will not be available under any agreement.

The findings from the survey have led to the conclusion that while much of the non-conventional literature generated in federal departments is collected now, the abstracting and indexing process to establish a new space data base is absent. Interest in participating varies from great enthusiasm to complete indifference, although no-one suggested that they would not make documents available. One private sector operator has submitted a bid to establish the energy data base, and expressed a wish to be offered the opportunity to do the same for the space

data base.

Based of the findings and the resulting conclusions, the following recommendations are presented.

- RECOMMENDATION 1.** The federal government, through the appropriate agency, should take steps to establish a Space Users Data Base to provide on-line access to non-conventional literature describing Canadian activities in the space field.
- RECOMMENDATION 2.** Since the new data base will start with documents, the format for storage and retrieval should conform to NASA standards.
- RECOMMENDATION 3.** The development of the data base should be placed in the hands of the private sector, but overall responsibility for regulation and policy should be retained by the federal government. CISTI is the logical candidate.
- RECOMMENDATION 4.** The data base should be based on current technical literature, with historical and scientific data entered as a second priority.
- RECOMMENDATION 5.** Establishment of the data base should be used in due course to negotiate an agreement with NASA for on-line access to its data base.

1. INTRODUCTION

1.1 Background

A number of surveys have been taken over the last few years to obtain an assessment of the state of awareness within the scientific and technical community of the opportunities presented by access to microgravity using the shuttle and, in the future, space station. The results have shown a wide divergence between the few who have been involved with space over the years and the many who are just now considering the possibilities.

A frequently asked question dealt with how to learn what information is currently available on the results of experiments that have already been performed. The purpose of the present study is to determine the feasibility of establishing a Canadian Space Users Data Base in order to provide that information. A key consideration is access to the NASA Aerospace Database where most of the international data reside. NASA has placed conditions on access; only those who contribute their own data will be permitted to obtain information on the work of others.

The premise is that by establishing a Canadian data base, and making that available to NASA, Canadians will gain access to the international information. NASA has already negotiated agreements with ESA and Israel for the exchange of information; discussions that have taken place between officials from CISTI and NASA lead to the conclusion that an agreement can be reached with Canada also.

1.1.1 A Canadian Data Base

The space science community has been active in Canada for many years. The National Research Council of Canada (NRC) has provided the leadership, in latter years through the Canada Centre for Space Science (CCSS). The program of CCSS has

provided funding for fabrication of launch vehicles in Canadian industry, including the installation of the instrumentation, and the overall management of the program. Funding is also provided for the purchase of launch facilities from other countries.

Investigators in universities have been supported by the Natural Sciences and Engineering Research Council (NSERC), and in government by the respective departments. The program is successful, and despite the recent cancellation of the long-standing balloon and rocket program, research in astronomy and aeronomy continues, and Canadian scientists enjoy a respected international reputation in this field.

The results of the scientific investigations appear in refereed scientific journals and are thus available through the normal channels; those that are "in scope" are included in the NASA Aerospace Database. There may be, however, internal technical reports that are not published in the open literature, and it is this non-conventional information that is of interest to NASA, and possibly to those who would be users of a Canadian Space Data Base.

The Canada Institute for Scientific and Technical Information (CISTI) carries a number of data bases in its On Line Enquiry system (CAN/OLE) that are devoted to the technical literature, as opposed to results that appear in the traditional scientific journals. Each of these data bases is maintained by the agency with operating responsibility in the subject area and is provided to CISTI for dissemination. A Canadian Space Data Base would fit this pattern, and if the requirement exists, it could also include references to results of scientific investigation.

1.1.2 The NASA Aerospace Database

The NASA file has two components. The International Aerospace

Abstracts (IAA) cover the world's published literature in the field of aeronautics and space science and technology. The Scientific and Technical Aerospace Reports (STAR) contain abstracts and indexes of report literature. The former can be compiled from the open literature, but the latter is more difficult to ferret out.

The many-faceted interests of NASA require a broad-based information bank with wide coverage and careful selection of reports, journal articles, books and conference papers. NASA's interests in science include

- environment and properties of the Earth, Moon and Planets
- the Sun and its relationships to the Earth and the rest of the solar system
- the space environment
- the physical nature of the universe
- the search for extraterrestrial life

NASA's interests in technology include

- spacecraft and launch vehicles
- aircraft including V/STOL, supersonic, hypersonic and lighter-than-air
- propulsion, auxiliary power
- human factors
- electronics
- structures and materials

NASA's interests in applications include

- astronomical, geophysical, meteorological and communications systems
- earth resources
- air and water pollution
- urban transportation

The scope of coverage can be judged by the general break-down of the subject matter contained in the overall data base.

aeronautics	9%
astronautics	5
chemistry & materials	11
engineering	20

geo-sciences	25
life sciences	4
math & computers	9
physics	10
space sciences	4
social science,mgmt.	3

The services are arranged in two sections, an abstracts section and an index section. The abstracts section has 75 subject categories and contains complete bibliographic citations with informative abstracts as required. The index section has five indexes

- subject
- personal author
- contract number
- meeting paper and report number
- accession number

Three levels of interest in each subject category have been identified

- exhaustive
- selective
- negative

NASA has suggested the following as potential users of the technology for civilian applications

- non-aerospace industries
- government at all levels
- educational institutions
- the medical profession
- non-profit organizations

Although the NASA coverage is immense (estimated listings 2.5 million growing by 125,000 annually), it should be remarked that much contained in the file can also be found in other directly accessible data bases. The distinction lies in the way in which the abstracting is carried out. NASA has a Thesaurus (NASA SP-7051) used for abstracting which is geared to the agency interest - space. There may therefore be some difference in abstracting emphasis between the same material contained in the NASA file and

other files. Furthermore, it appears that hard copies of the STAR abstracts can be purchased without restriction; there is a delay of about two years between publication date of the original document and the appearance of the record in the data base. Much of the information also finds its way eventually into the National Technical Information Service (NTIS), and can be obtained from that source in the usual manner.

2. METHODOLOGY

Since the most demanding aspect of establishing a Canadian data base will involve the collection of the non-conventional literature - technical reports, unpublished internal reports, reports that are not published in the normal scientific or technical literature - this feasibility study is concerned with ways and means of dealing with that issue.

At the present time CISTI carries a number of data bases the contents of which appear to match the NASA subjects. The main objective is to construct a Canadian data base, with the expectation that it can be used to negotiate access to the NASA Aerospace Database. If the necessary information is already resident in CISTI data bases, a review of the contents may reveal an efficient and effective means of satisfying the requirement.

Accordingly, the contents of a number of CISTI data bases covering the appropriate subjects were examined. This suggested that some information of the type sought is in fact, available through current data bases. It then becomes a question of how best to extract that information and recast it in a form suitable as an offering to NASA.

To answer this question, interviews were conducted with the appropriate data base managers to obtain an understanding of what would be required at what cost to take their present information and reformat it to conform to NASA standards. An interview guide was developed so that all necessary information would be obtained, and which would provide enough detail to allow an analysis to be made. The guide was refined in the first instance in consultation with information specialists from CISTI. In addition, a short background paper was prepared so that participants in the interviews would understand the nature of the undertaking. These are included in APPENDIX 1.

There is experience resident in Canada on the subject of contribution to international information services where there must be strict adherence to a given format. Atomic Energy of Canada Limited (AECL) at Chalk River has been providing abstracted and annotated material to the International Atomic Energy Agency (IAEA) in Vienna for many years. Since that experience would be useful in the present study, the proposed approach was tested with their specialists and modified accordingly.

The manager of the Sport Database is negotiating at present to convert to an internationally accepted system, so that it was considered useful to learn what conditions and problems were arising in that subject area.

Prospective interviewees were contacted to ascertain their willingness to participate, and the interviews were scheduled. The background document and the interview guide were sent in advance so that any detailed information requested could be prepared before the meeting. The list of those interviewed is to be found in APPENDIX 2.

The information obtained was collated and analyzed, which in turn led to conclusions and a set of recommendations.

We also reviewed a copy of the recently signed information exchange agreement between Israel and the US. This document illuminates the path that Canada may have to follow if we wish to enter into a comparable agreement.

3. FINDINGS

3.1 General

The experience of AECL in dealing with the International Atomic Energy Agency (IAEA) in Vienna was revealing and instructive. The experience in that field will probably guide events with NASA in respect of the Space Data Base. Negotiations relating to the Sport Database are in train now, and the situation can be monitored to see what can be learnt.

Although the premise that much of the "gray" technical literature that would form the basis of the Canadian Space Users Data Base is either obtainable or already being collected was found to be valid in principle, the state of development of many data bases is such that they can only be considered useful as a source of documents, or as entry points to literature in their particular field. The information collected and entered in machine readable form varies widely, from a bare minimum such as titles, authors and funding, to a complete abstracting and indexing service. Resources available to enter information in the data base are very limited in government departments, but the situation is different in the private sector where there are commercial opportunities. We learned that Energy, Mines and Resources has raised a requirement, through an RFP issued by DSS, to establish a Canadian Energy Data Base that would, in addition to holding Canadian information in one location, be used as a basis of an agreement with the US Department of Energy to permit Canadian access to the DOE Data Base.

We also learned that the STAR abstracts are available on a subscription basis. Since the IAA information is also available in the open literature, it appears that with much of the technical literature eventually reaching the NTIS, and the ability to purchase STAR abstracts, access to the NASA Aerospace

Database is not the sole means of obtaining information on space activities. However, ease of carrying out complex searches, efficiency and timeliness of information may rule out the subscription option.

3.2 INIS and Sport Database

3.2.1 INIS

The INIS data base is carried on CAN/OLE and covers world wide literature in the nuclear energy field. Access to this data base is made possible through agreement with the International Atomic Energy Agency in Vienna, and has been entered into on the understanding that Canada will provide INIS with Canadian non-conventional literature, that is, literature not available from commercial publishers.

AECL collects the information, and provides it to INIS in the form of machine-readable tape. There are about 600 entries per year, with most coming from Chalk River and Whiteshell. The abstracting is also done at these centres. The INIS file is not compatible with the CISTI format so that a conversion program must be used to convert to the CAN/OLE format. Each participating country has a liaison officer who provides INIS with country data and in return obtains the whole INIS file. There are no restrictions placed by IAEA on distribution within a country.

INIS does not retain hard copies of documents; the medium is microfiche. AECL provides hard copy of the documents it collects to CISTI, but retains only fiche for its own use.

AECL have 3 part-time abstracters, 2 at Chalk River and 1 at Whiteshell. There is also a clerical inputer and a descriptive cataloguer. The estimated the cost per entry is \$50, which

includes a copy of the document.

The following points were made.

- a) AECL believe that the operation of a data base such as contemplated for space should be centralized in order to ensure quality, continuity and that dedicated people are on the job.
- b) There should be one organization through which access to data bases requiring international agreement is channelled.
- c) The market may not be large, so that government support (possibly through CISTI) may be necessary to establish the initial data base in order to be sure it is set up properly.
- d) AECL information is carried on a CDC computer, which means that a translation routine is necessary to make the format conform to the IBM system in Vienna, which in turn must be translated to the NRC format.

3.2.2 Sport Database

The database was started in 1973 as an information service to Canadian sport coaches. Within a short time their mandate was expanded to cover all of sport in Canada. They started on a small scale and did not start to create their database until 1975. It took about 5 years for them to build the database to a size that could be distributed to users. Their file now contains over 200,000 records and is being added to at a rate of 15 to 18 thousand records per year.

The SPORT database has been recognized as the international database for sport by the International Association for Sport Information. They are also recognized as the sport database by UNESCO. This recognition has enabled SPORT to negotiate exchange agreements within 18 countries so far including the USSR. These agreements cover the contribution of data by these countries. In exchange SPORT does the indexing of each country's data and sends it back to the country as a file in machine-readable form. The country can then distribute the file in any way they see fit.

The agreements so far only cover conventional literature. In future it is hoped to expand the agreements to the non-conventional literature. They have tried to expand into non-conventional literature in Australia but have had very limited success so far.

SPORT is accessed through a number of database vendors which include CISTI, BRS (U.S. vendor). SDC, and as of January 1987, they are up on DIALOG. They are also now available through a West German vendor (since July 1986). They have agreements with these vendors to provide them the SPORT file. The vendor agrees to be responsible for his own conversion so the file can run on his system. The vendor pays a royalty to SPORT as a percentage of the on-line revenue plus a nominal amount for indexes cited. SPORT is responsible for maintaining and updating their file. Until April 1985, SPORT was part of the Coaching Association of Canada. They are now incorporated as a non-profit, independent organization. However 80% of their funding comes from SPORT CANADA. The remainder comes from royalties from the vendors.

SPORT has generated their own Thesaurus for use when indexing. They do not do any abstracting, relying mainly on the abstracts generated with the material. They use the abstract with the permission of the authors. The working language of the data base is English as well as the contributed abstracts. The original document can be in any language. They do not translate any original material except for the abstract if required.

Sources include journals, magazines and periodicals, and the coverage is close to 100% of sports-related literature in Canada. Canadian dissertations and theses are obtained, as well as some from the US. Cost is not a factor in collecting articles; hard copy is retained at their offices in Ottawa. The file is updated monthly.

The cost per entry is about \$15, which includes the salaries of the indexers, computing costs and acquisition costs. No capital costs are included and there is no provision for capital depreciation. The data base has cost about \$2 million to build to its present state, and much of the funding has come from SPORT CANADA.

The manager of this data base has considerable experience with international suppliers of data, and in fact, has proposed that his data base be managed internationally by the International Association for Sport Information. The reason is to reduce the cost of indexing and operating the data base so that resources can be made available to service the users through information dissemination. There is however, little information relevant to a space data base.

3.3 The Surveyed Data Bases

3.3.1 CODOC

CODOC is a catalog system developed at the University of Guelph to simplify and speed up the processing of government documents. It began about 20 years ago and is now supported by 10 academic libraries in Canada, a few in the U. S. and one in Saudi Arabia. Each library sends a tape of holdings to CrownTech in Toronto, where they are merged as a batch file. The merged tape is used to produce a microfiche record, and is then sent to NRC for loading onto CAN/OLE. The data base is also available for purchase.

The subject fields relate to primary interests of various faculties; at Guelph these are mainly agriculture, veterinary science and general science. Any subject matter to a space data base would only be by coincidence. Other universities may, of course, have some material that is more relevant.

Sources include any level and department of government.

Documents processed originate as follows:

Canada	30%
U. S.	30%
United Nations	30%
other	10%

At Guelph, a full depository is maintained for the Ontario government, with a selected collection from other provinces. For key areas of interest they have negotiated arrangements whereby they acquire one copy of each document published.

The CODOC collection could include consultants' reports, but nothing unpublished or "unconventional". Although Guelph does acquire some unpublished material, this is archived and subject to a 10 or 20 year time-to-release constraint.

All documents are held by and at the contributing libraries; there is no centralized holding area, and CAN/OLE receives only a tape. There are no rules for acquisition; each participant follows its own guidelines.

The catalog procedure involves about 20 separate data elements, all of a bibliographic nature. There is no abstracting or subject keywording; searches are carried out by matching words in the title. The file is updated bi-monthly, with about 15,000 to 20,000 new entries added. Holdings now number about 1.1 million. Costs are estimated to be in the range \$5-10 per entry, and about \$10,000 per year to mount the various tapes into CODOC, produce fiche and the NRC copy. About 4 person years of clerical staff is required to maintain the Guelph contributions.

There is limited compatibility with NASA requirements, and conversion to that format is considered to defeat the quick and easy access that is the feature of CODOC. There is really no interest in using or contributing to a space data base, at least on the part of Guelph.

The hardware used is an IBM compatible GEAC 8000 mainframe. The data base management system, which is MARC compatible, was developed at Guelph for GEAC machines. That company now markets the system, and Guelph receives royalties.

There is no interest at Guelph in accessing the NASA system, nor for that matter contributing to it, since the fields of interest have very little if anything in common.

3.3.2 CISTI Data Bases

Four data bases constructed and managed by CISTI were included in this survey. They are described separately below.

3.3.2.1 CISTI Catalogue Online (OON)

The OON data base contains records of conference proceedings, monographs, technical reports in science, technology and medicine held by CISTI, and union list records. The subject matter collected covers all the specified NASA areas of interest, although social science may be an exception.

CISTI maintains agents around the world who make acquisition arrangements through publishers and other sources. There are many standing orders in place with such agents. Most of the unconventional literature comes through exchange agreements with other agencies and institutions, but it was noted that branch libraries would be closer to the front line users and therefore closer to the non-conventional sources. All material collected by the branch libraries passes through CISTI Cataloguing and is recorded in the OON data base.

Documents acquired are stored, sometimes as original reprints, often copies, and some on microfiche, as for example, the NTIS documents. Cataloguing is continuous, and the file is updated on

CAN/OLE every two weeks. About 35,000 documents are added annually.

The entry system is based on MARC (Machine Readable Catalog), originally established by the Library of Congress, adapted to meet the DOBIS specifications. The final standard requires that the catalogue entry conform to a CISTI/DOBIS manual of practice which "distills" information in terms of CISTI policy. Staff must be highly trained - two years - to become an efficient cataloguer and to use this system. Unfortunately, the Library of Congress subject headings are not particularly strong in the sciences, and are seldom suitable for the very specific subject matter most often dealt with in technical reports.

A staff of 28 make up the cataloguing department, about half of which are librarians and half are clerical. Clerical staff handle most matters not dealing with original documents, and are not involved in original cataloguing or creating new records.

Without a detailed review of the NASA format, it was not felt that the closeness of the match could be well defined. To meet NASA standards a major cost would be incurred in abstracting. This was estimated to be about \$50 per document.

3.3.2.2 IEC

The IEC data base was started in 1971/72 to track federally funded university research. The data base does not reference literature, rather it details funding allocations. Coding is done within IEC by a private contractor, Data Cap of Ottawa. The contractor is given the original documents, such as annual reports, from which they must extract information. Records are entered twice - by different people - and the results matched for verification.

The largest source of information is NSERC, who provide grants data on magnetic tape. IEC has a conversion program to pick off and manipulate appropriate information and assign it to relevant fields. Government departments also supply information.

Coding includes project title, investigator(s) along with University and department. Financial information is also provided. No abstracting is performed, and searches are executed by matching words or names in title fields.

Approximately 15000 records are added annually, at a cost of one person year of effort. An annual catalogue is published listing all the data and definitions with cumulative totals provided by subject. This document is more frequently consulted than is the on-line data base.

3.3.2.3 NRC/PUBS

NRC/PUBS lists publications produced by and for the NRC. The data base was established in 1983. At that time there were approximately 20,000 listings in existence, about half of which were contained in various machine-readable formats. System House was retained to write conversion programs, and the 10,000 listings thus entered formed the original file.

Those records for which conversion programs were written contained only elementary bibliographic data stored mainly as free text. Fortunately the text in particular record groups generally followed patterns which the computer could be programmed to recognize. Some manual correcting was required and this took about one year to accomplish. The remaining 10,000 listings are entered manually on a low priority basis. NRC/PUBS is compiled on MINISIS.

An archive of almost all material is kept by NRC/PUBS. Original reprints are preferred to copies, and some items have been

microfiched. This archive is a backup for the CISTI document retrieval system.

Since the file contains NRC data, some subject fields will match the NASA requirements. However, the format in which the data is stored may make it difficult to construct a translation routine that could extract the desired information automatically. The basic documents are certainly available.

NRC/PUBS is maintained by a staff accounting for about 1-2/3 person years per year.

3.3.2.4 Translations Index

The Translation Index (TI) is a listing of documents which have been translated into English and/or French. It includes data from mainly international sources with very little Canadian material. The major sources of material for TI are government departments and agencies who have their own networks for collecting material for translation. The translations are performed in these agencies, and the listing sent to TI for entry in its data base.

Information is entered with bibliographic references only. There are no abstracts, subject headings or keywords; searches rely on name (first author only) and title word matches. Only about 1/3 of all translations listed are obtained and stored. Documents for 50,000 entries are currently held. Originals are kept at present but microfiche will be employed in the near future.

TI tries to collect as many STI documents as possible; however there are very few closely related to space. A quick search using the keyword "space" revealed a match of only 0.1%. The TI file is maintained using MINISIS software developed by the International Development Research Centre (IDRC). MINISIS is

based on the mainframe system ISIS and runs on the HP 3000 mini-computer.

An entirely new format would be required to produce NASA compatible records. In addition, abstracting experts would have to be retained. At present there is a permanent staff of two expanded to 4 or 5 with temporary contracts.

**3.3.3 MICROLOG, Canadian Business and Current Affairs,
Directory of Associations in Canada**

These three data bases are managed by Micromedia Limited of Toronto, and are offered on a commercial basis. Canadian Business and Current Affairs (CBCA) covers news items from 10 Canadian newspapers; editorials; selected letters to the editor; and other items of general interest contained in newspapers. It also covers 170 business periodicals. No abstracts are available. The Directory of Associations in Canada (DAC) contains addresses and other information for active Canadian associations. Abstracts are available. Neither of these data bases contain primary information of interest to a space data base, but they may point the way to sources of interest.

MICROLOG is a different story. The subject content is multidisciplinary including technical and annual reports, research studies, statistical reviews and financial statements. Through a wide and effective network of information sources, MICROLOG acquires the non-conventional literature (45% STI) from Canadian federal, provincial and local governments, non-government research institutions, professional associations and special interest groups. Some of that information may be reformatted to conform to the NASA requirements; more important is the network that exists which can be used to collect Canadian non-conventional literature.

Micromedia currently contribute to about 20 on-line systems across Canada. They are also the sole agents for the US DIALOG system, maintaining training, customer support and marketing operations in this country. As a result, they are familiar with the content and structure of each DIALOG file they service. Micromedia have a staff of about 65 full-time people plus about 65 people on contract, allocated to functional areas. Of the full-time employees, 15 are information science/librarians, 15 are technicians (fiche etc.), 15 are library technicians, and about 10 in management. The MICROLOG data base is directly supported by 5 full-time people, two acquisition people, one cataloguer, an administrator and an abstractor. The librarians have national contacts, and this forms a significant part of the information network, which is informal but reaches across the country.

When a document is acquired, some tombstone data such as author, date, institution etc. is added, and then the indexing and abstracting takes place (in both official languages for STI) - simultaneously. Documents are not held, but are microfiched for storage and the provision of hard copy. Micromedia have reached an agreement with CANMET, EMR, to carry their technical literature, which includes producing microfiche. This was held to be a good example of how the private sector can work with government to optimize the STI process. Micromedia also have the rights to the patent information from Consumer and Corporate Affairs, and in fact have an individual located in an office of that department provide liaison. They have also received a contract from the National Library to film dissertations. Overall it is estimated that the ratio of gray/non-gray material held is about 3/2.

The steps that are involved in their business were identified as

- acquisition
- cataloguing

- indexing/abstracting
- production of machine readable tape (9 track Mag.)

The hardware is an Alpha-micro (actually a mini) linked with 25 terminals, using the Alpha Micro Operating System (AMOS). The software has been developed in-house, but some use has been made of the STAR data base creation system (not to be confused with the STAR abstracts) produced by Quadra Associates. An important feature of STAR is that it incorporates a real-time error-checking facility.

About 30 abstracts per day are produced for MICROLOG at a cost of about \$50/abstract. The file is updated continuously and about 7000 entries are added per year. DAC lists 12,000 associations; CBCA lists 60,000 articles per year.

As the Canadian outlet for DIALOG, Micromedia have a good sense of what would be required to develop a NASA-compatible data base. They view the establishment of a space data base as a commercial opportunity and expressed an interest in developing such a data base. They did, however, recognize the need for a federal agency to deal with policy and regulatory issues.

3.3.4 GEOSCAN

GEOSCAN developed from a proposal made to a Mines Ministers' Conference in the late 1960's. It is organized as a federal/provincial cooperative, with 3 full-time employees in the Geological Survey of Canada coordinating the activities of 15 indexers working for provincial agencies across the country.

Only a small percentage of NASA subject matter is covered in the file. Furthermore, most holdings fall within the NASA "negative interest" category. The primary sources of information are

- mineral exploration reports
- government reports
- open file reports (work in progress)
- journal articles

About 65% of the reports are unpublished, and current holdings include some 110,000 documents, 59,000 of which are exploration assessment reports. Some National Topographic System (NTS) maps are also included. Documents are held by individual provincial agencies; there is no central archival or distribution agency. Quantity of documents acquired depends on the resources of the individual participating agency. About 5000 entries are added each year, but no abstracting is performed at present. The cost is estimated to be about \$100 per entry. The file is not carried on CAN/OLE.

3.3.5 Canadian Transportation Documentation System (OOT)

Information collected by this system is restricted to that which is related to the work of the department. All departmental published reports and books, with the very occasional journal article are held, and an attempt is made to obtain consultants reports. Both commercial and government agencies are sourced for literature. Transport Canada issues Transportation Publication (TP) numbers for all their publications. These numbers are checked to ensure that all documents are obtained and listed.

Other sources include agreements with outside organizations; in the past all Statistics Canada documents, and many NATO, international marine organizations, and DSS documents have been obtained through standing orders for related subject matter. The NTIS is a major source of documents. Reports produced by the Transportation Development Centre (TDC) are obtained and listed. The Roads and Transportation Association of Canada is a good source.

Documents are held in the head library in Ottawa (75%) and branch libraries across the country. The data base file is updated monthly. No abstracting is done because of the high (but unspecified) cost. Most TDC reports have an author generated

abstract, but the current format has no provision for storing them. About 8000 entries are made per year.

It is estimated that there would be very few holdings of interest to a space data base in the fields of aerospace, astronautics, chemistry and geo-science. Engineering as it relates to avionics, particularly IEEE publications, would be fairly well stocked.

Cost of document handling and storage is not available, although about 10 years ago it was estimated that records cost \$5-6 each, with no certainty what this included. About 25 documents are catalogued daily for the expenditure of \$75,000 annually, not including branch library time.

In the present configuration the impact of providing data in the NASA format would be enormous. There is no expertise in abstracting. In terms of participating, there is a willingness to assist in locating documents, but not in cataloguing or abstracting.

OOT is created on an IBM mainframe maintained by Transport Canada using LIBIS, a version of the DOBIS system. There are plans to install a modification that may make it easier to translate to NASA standards should that become desirable.

3.3.6 Remote Sensing On-Line Retrieval System (RESORS)

The RESORS data base was started in 1972, and was based on the card catalogue of the CCRS library. The system has been custom built, originally by summer students, and the catalogue/index procedures were developed specifically for this application. It was established to meet the needs of the in-house scientific research staff.

There is some overlap between the fields of interest of NASA and those of RESORS, but the different focus in each case results in very little actual material of common interest. The majority of their holdings are of foreign origin, with Canadian material accounting for about 10%.

The primary sources for information are other data bases such as GEOREF, NTIS, the American Society of Geographers data base, and CISTI data bases that contain material of relevance. They also have standing offers with other agencies, and do carry out an annual search for theses on the US University Microfilm - an on-line retrieval system. They also enter data on remote sensing symposia.

All documents catalogued are held, and stored in the RESORS offices in either paper copy or microfiche. They also maintain a searchable file of some 6000 slides. Additions are appended to the file monthly.

No abstracting is performed, and searching is only possible by matching keywords that are entered into the file when the document is processed. Current holdings are about 50,000, with about 5,000 added annually. The format is not compatible at all with that of NASA.

Document processing is contracted out to Gregory Geoscience who maintain 5 full-time staff, of whom 3 are indexers, on the CCRS premises. The contract is worth \$143,000 for the current year, but computer time and storage space are provided by CCRS.

CCRS has some interest in accessing the NASA Aerospace Database and would be willing to provide information on their own activities. They have obtained some documents from NASA, but most have come from NTIS.

3.3.7 Department of National Defence (DSIS)

The Defence Scientific Information Service (DSIS) was established in the 1940's as part of the post-war Defence Research Board, and went on-line in 1969. Its purpose is to serve the Defence Department and its authorized contractors and university staff.

About 99% of the holdings are in the non-conventional category and originate in the Defence Department, other Canadian sources and foreign countries. Material from the Defence Department, amounting to about 25% of the total, is sent to the Defense Technical Information Centre in the US where it is screened for security. Information considered unclassified is released to the NTIS. The remaining 75% is retained in Canada.

Abstracting and indexing are carried out by 5 people, working about one quarter time. They work from a Thesaurus that was originally developed by the Department of Defense in the US but was never adopted by that country, on the grounds that there would always be changes, and that the Thesaurus should be dynamic! The result is that most NATO countries use the original Thesaurus as a base and build changes as time goes on - but the most dynamic is that used by the US. The situation makes standardization amongst countries impossible, although efforts are made continuously to arrive at a common Thesaurus.

The Canadian data base contains about 80,000 records with 5000 added annually. The file is updated continuously; one third to on half are of Canadian origin. Comprehensive abstracts are produced, sometimes as long as 1000 words, but there is no standard format. However, the system is such that the operators are unwilling to consider a change to meet NASA standards. Most hard copy, in microfiche or document form, is stored on site.

No cost information was available, but the figure of \$50 per

entry was considered reasonable.

DSIS is anxious to expand the data base to include information from a number of sources that do not contribute at present, mainly a number of other government departments. There is great interest in accessing the NASA Aerospace Database and unclassified information would be contributed on the understanding that access would be assured in return.

Three to four years ago DSIS made an effort to negotiate access to the NASA Aerospace Database, with the assistance of CISTI. A Mr. Thibideau from NASA visited DSIS to discuss the requirements, but the process was considered too cumbersome to pursue. DSIS does subscribe to the STAR abstracts, which are reasonably current, but the process of trying to extract information from this very extensive document is too time-consuming to be carried out on a routine basis. There are two classes of STAR information - classified and unclassified. Only the latter are available for purchase. DND interest in access to the NASA Aerospace Database is intense, unquestionably the greatest encountered in this survey.

3.4 Sourcing List - Department of Supply and Services

The Department of Supply and Services was contacted to ascertain whether the source list of suppliers for space-related work would be made available for the new data base. We were informed that such information would be made available, but due to the development of a new system within the department, sourcing information will not be available before October, 1987.

3.5 The Israel - US Agreement

In accordance with NASA principles, the basis for the agreement lies in the provision of domestic space-related information from

Israel in exchange for access to the NASA Aerospace Database. On the surface, this does not seem difficult to implement. However, our reading of the undertakings by Israel suggests to us there is a requirement is that the lead Israel agency must undertake to enter into a back-to-back agreement with each institute within that country who wish access to the NASA file. This would seem to limit access to members of the particular institutes who are willing to make the effort to collect and provide information for onward transmission to NASA.

While it is our understanding that each agreement signed by NASA may have different conditions attached, nevertheless the widely-distributed nature of the Canadian information will make it difficult to develop back-to-back agreements and still achieve the main purpose of an agreement with NASA - access by Canadians to space information.

3.6 Sample Searches

We tested a few systems to see whether in fact they do contain documents we know would be included in a space data base. MICROLO appears to have an extensive coverage of the type of technical literature sought, so a search was carried out to see if the Proceedings of a Workshop on Materials Processing in Space could be retrieved; it could not. The keywords "space", "microgravity", "medicine" and "life sciences" were then used in combinations to see if any documents with those descriptors could be identified; again the results were negative.

Our search of OON revealed about 3000 entries responding to the keyword "space", which can encompass more than the subject matter of this study. As noted in a previous section, a search of the Translation Index using the keyword "space" revealed a match of only 0.1%.

We also tried a variety of combinations on OON, as we did on MICROLOG, and the number of hits were always less than 10.

3.7 Summary

A considerable amount of diverse information about the data bases surveyed has been collected, and in order to provide a synoptic overview this has been tabulated in Table 1 on the following pages. The information is presented in the general format of the interview guide, so that where specific data were collected, they can be so displayed. The table provides a quick summary of information contained in the sections immediately above.

Table 1
SUMMARY OF FINDINGS

	CODOC	IEC	MICROLOG CBCA DAC	OOB	GE
Related fields covered	Government Science Transportation	Chemistry Computer Science Engineering Geology Medicine Physics Science	Behavioural & Social Sciences Communications Life Sciences Natural Resources & Earth Sciences Physical Sciences Transportation Can.Bus.Index Can.News Index Associations	Aeronautics Astronomy Biochemistry Chem. Engineering Chemistry Computer Science Elect.Engineering Information Science Materials Science Mathematics Mech. Engineering Medical Sciences Physics	Geo-sci Math/com related
Estimate of percent covered	uncertain, but probably little	n/a; lists funding	need to do a search	all, except maybe some social sci.	only small percentages
What not covered and why	subject areas related to faculty interests	n/a	as above	no restriction	not within purview
Sources	departments of governments	mainly NSERC	not driven by cost question of locating sources	exchange agreements branch libraries	exploratory government open files
Documents located	contributing libraries	no documents	held in fiche form	fiche, reprints copies - M55	individual provincial
What determines cut-off	no rules	subject matter funded by grants	supply	funding limits	discretionary individual
Updated	bi-monthly	annually	continuously	every two weeks	not known
Data elements employed	many, all of a bibliographic nature	title investigator university department amount funding agency fiscal year funded language		hundreds	author geologist geographical NTS keywords
Abstracts/titles	title keyword only	project names	abstracts	no-headings, keyword	no abstracts

Table 1
SUMMARY OF FINDINGS

OOH	GBOSCAN	OOT	DSIS	RESORS	INIS	SPORT
Aeronautics Astronomy Biochemistry Chem. Engineering Chemistry Computer Science Elect. Engineering Information Science Materials Science Mathematics Mech. Engineering Medical Sciences Physics	Geo-sciences Math/computers related to mapping	Surface Marine Air Off-highway Intermodal	Aeronautics Chem. & Materials Engineering Human performance Math & Computers Physics Soc. sci. & mgmt. Defence related	Applications Photogrammetry Image analysis Aeronautics Astronautics Engineering Geo-sciences Space Sciences Physics (optics)	All topics related to nuclear energy; ARCL collects and provides all Canadian information	individual & team sports practice techniq. int'l sport hist. phys ed recreation facilities sport medicine training & equip.
all, except maybe some social sci.	only small percentage	does not use subject headings	large percentage	very little of Can. literature	90% of Canadian	none
no restriction	not within the purview	aeronautics chemistry/aerospace geo-sciences limited funds	nothing specifically excluded	limited interest	no topics excluded	n/a
exchange agreements branch libraries	exploration reports government reports open file/journal	commercial and government agreements, NTIS	DND, some OGD's, Foreign (about 75%)	data bases US Univ. Microfilm standing offers	ARCL Gov't depts Can. Nucl. Assn	journals other countries dissertations
fiche, reprints copies - M55	individual provincial agencies	Ottawa 65% branch libraries 25	DSIS	RESORS offices	Can. at CISTI INIS in Vienna	River Rd. facility
funding limits	discretion of individual agencies	tries to be exhaustive	tries to be exhaustive	resources	resources	availability of material
every two weeks	not known	monthly	continuously	monthly	ARCL 2mo.	monthly
hundreds	author geolog formations geographic areas NTIS keywords	many	many	accession number author title keywords		keywords from their own Thesaurus
no-headings, keyword	no abstracts	no abstracts	abstracts	no abstracts	abstracts	no abstracts

ODN	GEOSCAN	DOT	DSIS	REGORS	INIS	SPORT
35000/year	4-5000/year	1000/no. delations result in 8000/yr	5000/year	5000/year	AECL 600/yr INIS 70,000/yr	15-18000/yr
uncertain	uncertain	NASA much more detailed	no	not at all	very little compatibility, but may be irrelevant	very little compatibility
changing MARC would be difficult	system has memory limitations	no expertise in abstracting	would have to start from the beginning	total reworking of the system	may have to start from scratch	need to be re-indexed
DOBIS, on an IBM 309200, using MVS operating system	MINISIS on an HP 3000	LIBIS, a version of DOBIS, on an IBM mainframe	specially developed	PDP 10	screen editor, CDC computer	MINISIS
not known	\$33	not known	not known	about \$4/doc.	no docs at AECL	not known
not known	\$100	\$5-6	not known	about \$30	\$50	\$15
not known	\$200k/yr to System House	not known	about \$60,000	\$143,000	\$40K	about \$225,000
n/a	no	not known	n/a	no	no	no
n/a	no	not known	n/a	no	no	no
not known	not known	not known	n/a	not known	n/a	not known
yes	no	yes	yes	yes	some indirect	yes
NRC-responsibility	n/a	department mandate	department mandate	department mandate	Can, CISTI, AECL	Sport Can. mandate
14	about 7	3	-	3	-	3
-	-	-	8	-	3	-
-	-	-	-	-	-	-
about \$50	uncertain	not known	not known	not known	probably \$50	not known
need more staff if there will be more work	impact on provincial agencies who have to do the work	traumatic; no expertise	would have to start abstracting process from the beginning	would have to start from the beginning	very little in INIS that would be of interest to NASA	not much of interest

	CODOC	IEC	MICROLOG CBCA DAC	ODN	GI
Titles and Entries per year	about 100,000	15000	7000 in microlog	35000/year	4-5000/y
Conform to NASA data elements - format	limited	no	have experience in matching, don't anticipate trouble	uncertain	uncertain
Problems converting to NASA	better to start afresh; CODOC is "quick", this is its advantage		none	changing MARC would be difficult	system / limitation
DBMA/hardware	MARC format compatible; Guelph developed; IBM	NRC mainframe and software	internally developed (STAR) Alpha micro	DOBIS, on an IBM 309200, using MVS operating system	MINISIS 3000
Cost/doc to handle	n/a	about \$2-3		not known	\$33
Cost per entry	about \$5-10	n/a		not known	\$100
Annual operating cost	about \$10,000	1 py	n/a	not known	\$200k/yr House
Deprec. included	no	no	n/a	n/a	no
Cap.cost included	n/a	no	n/a	n/a	no
Capital cost	n/a	n/a	n/a	not known	not known
Subsidy	some	yes	no	yes	no
Who/why	university shared	NRC- responsibility	n/a	NRC-responsibility	n/a
No. abstractors. FT PT contract	4 0 0	1 0 0	15 - 15	14 - -	about 7 - -
Cost per entry to meet NASA standards	not known		about \$50	about \$50	uncertain
Impact on operation to meet NASA standards	could not set up required system as they hold so few in-scope documents		greater business opportunity, if they provide the data base	need more staff if there will be more work	impact providing who have work

DOH	GEOSCAN	OOT	DSIS	RESORS	INIS	SPORT
known	casual	limited; STAR and IAA documents are eventually available	very interested; looked into the possibility some years ago	some interest; currently use NTIS	very little	very little
	subject to prov. agreements	no abstracting or cataloging	yes-would provide documents	yes-preprints	probably	yes-would provide documents
staff	uncertain	n/a	require access to NASA Aerospace Database	none	if AECL becomes involved in space; if funds provided; if management agrees	would not index or abstract
known	uncertain	not identified	save much time	would enlarge scope of RESORS	not known	few if any
ertain	uncertain	space agency for federal role; private sector for the service	requires federal leadership, at least in early stages	separate organization	centralize effort and people	none
ertain	uncertain	no comment	no comment	no comment	likely, as market is limited	not known
	uncertain	no comment	no comment	as per their own experience	STAR abstracts available in hard copy; AECL buy DOE energy data base;	might be possible to give documents to NASA, let them do the abstracting, and get access
adian content out 15-20%; such libraries y be better nacts	may be a move to rationalize all government data bases into one data base	installing a new system that might lead to a cataloging capability that would be closer to NASA requirement	intend to extend scope of their coverage to include more OGD's, NSERC, other university support, Science Council	none		might be better to give docs to NASA and let them do the abstracting in return for access

	CODOC	IEC	MICROLOG CBCA DAC	OOH	GROS
Interest in using data base	none	none	only as a commercial information supplier	unknown	casual
Willing to cooperate	no; would require a change in methods		yes	yes	subject to agreements
Caveats			want to develop the system on behalf of sponsoring agency	more staff	uncertain
Benefits to organization	none	none	increased scope of operations	unknown	uncertain
Preferred organizational structure	none	no opinion	private sector to create and run DB, public sector policy and regulation	uncertain	uncertain
Federal role	no opinion	no opinion	policy and regulation, funding	uncertain	uncertain
Commercial opportunity	no opinion	none	yes, if private sector develops and runs the data base	no	uncertain
Comments	none	annual catalogue published	would like to set up the data base; bidding on similar requirement for RMR have a well developed information net already in Canada	Canadian content about 15-20%; branch libraries may be better contacts	may be a rationalized government bases into base

4. CONCLUSIONS

The premise that led to this study, that a considerable amount of the Canadian non-conventional literature of interest in establishing a Canadian Space Users Data Base was already being collected through the acquisition policies of existing data base operators, would appear to be borne out only in a very general sense. In no case did we discover that there is a one-to-one correspondence between the fields of interest already covered and those anticipated for the new data base. In other words, the construction of translation routines would not yield a data base with the required attributes.

At the same time, a number of data bases appear to include data of interest, although not filed in the desired format. In many cases abstracts are not provided; data fields differ; the purposes in establishing each data base are different. However, we have learned that there are very extensive and active networks throughout the country already at work collecting information for existing data bases. Our judgement is that these networks could be used to obtain access to the data for the space data base.

The majority of the data bases surveyed are operated under the auspices of federal government departments. This limits the extent to which data can be collected to that necessary for the particular department to carry out its mandate. The exception is CISTI (NRC), which has responsibility for collecting and holding scientific and technical information without reference to a particular departmental program. The extent to which data are collected is then dependent on the availability of sources and of resources.

The sources for all government-operated data bases start within the department itself. There is a system in place to collect technical documents and to record their existence. The manner in

which the latter is done varies from data base to data base. In a few cases (INIS) there is an abstracting service; the majority follow library-type classification systems, Library of Congress or a variety thereof. This is adequate for the purposes for which the data bases were intended. In one case, OON, the basic requirement is for an efficient cataloguing system.

In all cases, the information is being updated at regular intervals, so that the data base is dynamic. The resources allocated to the activity (apart from CISTI), appear to be modest, and any extra demands placed on them to conform to the NASA standards, for instance, would create an enormous burden - one which we judge they would be unwilling to assume.

Cost information was not readily obtainable, and that which was forthcoming is of doubtful value in many case because the basis is uncertain. Certainly in the case of government departments, capital costs are considered as sunk costs and are therefore not included in any cost calculations. This is in contrast to the private sector data base operation we visited.

We received a range of estimates for the cost of entering information into a data base, starting with a document, which also must be retained. The realistic cost is about \$50/entry. Judging from the response to the question relating to the extent of NASA "in scope" coverage, it appears that only three of the surveyed data bases contain much that is relevant. These are OON, DSIS, and MICROLOG; entries amount to about 50,000 per year.

Based on our test searches, there is little evidence that many documents have been entered in the past which have been catalogued with our space interests in mind. However, that situation could change quickly with the establishment of a subject data base aimed at space users. It may be premature to give a reliable estimate of the number of entries per year, but

it is unlikely to exceed 600. Based on \$50 per entry, \$30,000 per year should cover costs for at least the first few years.

Hard copy is available for data contained in the CISTI data bases and also for those operated by the private sector. Storage is handled differently by different operators; CISTI retains documents and microfiche; Micromedia retain only microfiche. Both provide, and charge for, hard copy on request.

A number of different hardware and software systems are used. This does not generally pose an insurmountable problem in moving data between systems, even reformatting is required. Translation routines are used, and while these are often difficult and time-consuming to develop, once in place they work well. Upgrading is also taking place as new requirements are raised and/or new technology appears on the scene.

The wide-spread availability of Compact Disk (CD) technology has prompted Micromedia to examine the applicability of that medium to the storage and retrieval of information. The enormous storage capacity and the security of data assured by the read-only nature of the disk are attractive for some applications. There is an interesting problem for commercial operators; pricing strategy is a non-trivial matter. If the cost to the customer is too high, there will be few sales; if too low, other information services, which may in fact be the mainstay of the CD, will wither away. At the moment, it appears that the near-term configuration will evolve to a combination of CD and on-line. The former will handle data on say, an annual basis, and the latter the monthly add-ons. This implies that a new CD would be issued each year.

As mentioned previously, Micromedia is aware of the requirement to establish an energy data base. During our interview, we were told that they would be very interested in performing that task

for a space data base. In both cases they have access to documentation on the US data bases through their DIALOG affiliation. They believe they have a good assessment of the cost of establishing a data base of the type visualized for space users.

In answer to our questions regarding appropriate roles and organizational structures, there was unanimous agreement that the overall responsibility should rest with CISTI, an organization with an enviable reputation. At the same time it was noted that the energy data base will be run by EMR who will sign the agreement with the US Department of Energy. This may have implications for CISTI and the NRC when the proposed Space Agency comes into being. The private sector, Micromedia, would like to develop the data base and market it through CISTI as is now done in the case of the data bases referred to in our summary of that interview. There was complete acknowledgement that the agreement with NASA should be negotiated and managed by CISTI.

In summary, we conclude that the simple reformatting of technical literature held in the data bases surveyed will not suffice to form the basis of a Space Users Data Base. At the same time, with very little additional effort, the relevant documents can be collected through existing networks. The steps of abstracting/indexing, making fiche, entering data and managing the data base will require resources, estimated at about \$50 per entry.

Furthermore, we believe that the structures are in place that will be needed to establish a Canadian Space Users Data Base.

The Israeli agreement and its implications for Canada lead us to conclude that the primary objective of a Canadian Space Users Data Base should be to serve Canadians, and the approach to establishing such a data base should take place with that in

mind. It may turn out that the sources are sufficiently concentrated that back-to-back agreements make sense. The AECL experience is a case in point. On the other hand, while CISTI can deal with the federal end, it does not seem reasonable to try to enter into agreements with every university, institute or private sector company. We therefore conclude that the Canadian needs be satisfied first, and when the pattern of supply becomes clearer, negotiate an agreement with NASA.

5. **RECOMMENDATIONS**

This study has one objective - to determine the feasibility of establishing a Canadian Space User Data Base. However, this has not been considered in isolation from the prospect of using such a data base to gain access to the very much more extensive data base maintained by NASA. Therefore, a large part of our exploratory discussions with various data base managers has dealt with the issue of providing Canadian data in a format acceptable to NASA.

We have concluded that the establishment of such a data base is indeed feasible, that it must start with the acquisition of appropriate documents through information networks used by the managers of existing data bases, and that a separate process must be put in place to abstract, index, enter and store the resulting data. There is at least one private sector operator who has the willingness and capability to undertake such a task.

NASA documents are not available solely through the NASA Aerospace Database. The scientific literature is available, coded perhaps a bit differently, through the IAA. Abstracts of the technical component, STAR, are available as a commercial publication. Much of the technical material finds its way eventually into the NTIS. Although these alternatives exist, timeliness and ability to manipulate data may be factors, and on-line access may be considered essential.

In light of the above we submit the following.

RECOMMENDATION 1. **The federal government, through the appropriate agency, should take steps to establish a Space Users Data Base to provide on-line access to non-conventional literature describing Canadian activities in the space field.**

- RECOMMENDATION 2.** Since the new data base will start with documents, the format for storage and retrieval should conform to NASA standards.
- RECOMMENDATION 3.** The development of the data base should be placed in the hands of the private sector, but overall responsibility for regulation and policy should be retained by the federal government. CISTI is the logical candidate.
- RECOMMENDATION 4.** The data base should be based on current technical literature, with historical and scientific data entered as a second priority.
- RECOMMENDATION 5.** Establishment of the data base should be used in due course to negotiate an agreement with NASA for on-line access to its data base.

APPENDIX 1

Background Paper and Interview Guide

A Study to Determine the Feasibility
of
Establishing
a
Data Base
for the
Space Station User Development Program

Background

A decision has been taken that Canada will participate in the US Space Station Program, and part of this participation will involve using the space station as a platform for experimentation. Canada has been active in space for many years but there is an expectation that access to a platform in microgravity will lead to a significant increase in space R and D.

Other countries, particularly the US, have gained considerable experience working in space and are well up on the learning curve compared to Canada, in most fields. Catching up will not be easy, but access to the results obtained by others will speed the process. At present this is restricted, resulting from a US policy that only contributors can have access to data already obtained. This study is intended to assess the feasibility of developing a Canadian data base that can be provided to the US in return for access to the NASA Data Base.

NASA will negotiate agreements on a country-to-country basis and have done so for ESA and Israel already. Other countries are expected to follow. If Canada can also participate, we will automatically gain information about ESA and Israeli findings; also those of other countries as agreements are concluded.

The NASA file has two components. The International Aerospace Abstracts (IAA) cover the world's published literature in the field of aeronautics and space science and technology. The Scientific and Technical Aerospace Reports (STAR) contain abstracts and indexes of report literature. The former can be compiled from the open literature, but the latter may be more difficult to ferret out. This report literature will form the contents of the Canadian data base, and is the subject with which this study is concerned.

The many-faceted interests of NASA require a broad-based information bank with wide coverage and careful selection of reports, journal articles, books and conference papers. NASA's interests in science include

- environment and properties of the Earth, Moon and Planets
- the Sun and its relationships to the Earth and the rest of the solar system

- the space environment
- the physical nature of the universe
- the search for extraterrestrial life

NASA's interests in technology include

- spacecraft and launch vehicles
- aircraft including V/STOL, supersonic, hypersonic and lighter-than-air
- propulsion, auxiliary power
- human factors
- electronics
- structures and materials

NASA's interests in applications include

- astronomical, geophysical, meteorological and communications systems
- earth resources
- air and water pollution
- urban transportation

NASA has suggested the following as potential users of the technology for civilian applications

- non-aerospace industries
- government at all levels
- educational institutions
- the medical profession
- non-profit organizations

Three levels of interest in each subject category have been identified

- exhaustive
- selective
- negative

The services are arranged in two sections, an abstracts section and an index section. The abstract section has 75 subject categories and contains complete bibliographic citations with informative abstracts as required. The index section has five indexes

- subject
- personal author
- contract number
- meeting paper and report number
- accession number

The scope of coverage can be seen by the general break-down of the subject matter contained in the overall data base.

aeronautics	9 %
astronautics	5
chemistry & materials	11
engineering	20
geo-sciences	25
life sciences	4
math & computers	9
physics	10
space sciences	4
social science, mgmt.	3

Although the coverage is immense (estimated listings 2.5 million growing by 125,000 annually), it should be remarked that much contained in the NASA file can also be found in other directly accessible data bases. The distinction lies in the way in which the abstracting is carried out. NASA has a Thesaurus (NASA SP-7051) which is the Bible of its data base, and which is geared to the agency interest - space. There may therefore be some difference in emphasis between the same material contained in the NASA and other files.

Based on agreements already in place, Canada will probably have to undertake the following

- designate one agency to obtain and contribute in-scope material
- this agency will have to use best efforts to ensure that each participating Canadian organization makes its maximum in-scope contribution
- in return for contributions, the agency will recommend to NASA that those Canadian organizations be approved for commercial access to the "Aerospace Database"
- the agency will keep records of contributions and report to NASA every six months
- the agency will process the material and provide NASA with tape and microfiche
- the material to be provided includes the following
 1. In-scope technical reports, working papers, conference papers and proceedings
 2. Pre-prints of in-scope journal articles
 3. Space Agency information bulletins
 4. Directory of Canadian researchers
 5. In-scope university theses and dissertations
 6. Relevant STI publications from other government organizations

INTERVIEW GUIDE

I. Contents

1. What was the motivation to start.
What problems were encountered.
What process was followed.
2. Is the list of the subject fields covered on CAN/OLE data sheet correct. If not, what additions/deletions.
3. What percentage of the NASA subject matter do you estimate you cover relating to Canadian literature.
 4. If not all, which are not covered and why.
 5. If because of marginal costs, can you quantify.
6. What are the sources.
7. Where are documents kept.
8. How do you decide when you have acquired enough .
9. How often is the file updated.
10. What data elements (fields) are used.
Which if any are mandatory (ie. searchable index fields)
11. Abstracts or just titles.
12. How many new entries per year.
13. How close is the match with NASA.
 - a) data elements
 - b) format
14. What problems would you anticipate in converting to NASA standards.
15. What data base management system is used to create the file.
What hardware is employed.

II. Costs

16. What cost per document, to obtain, handle, store.
17. What is the cost per entry.
18. What is total annual cost of operating the system.
 19. Is capital depreciation included.

20. Is capital cost included.
21. What was capital cost.
22. Is there a subsidy. How much.
23. If yes, from whom and why.
24. How many abstractors. Full time, part time, contract.
25. What would be the additional cost per entry to meet NASA standards.
26. What would be the impact on your organization to produce records of in-scope material according to NASA standards.

III Future

27. What is your interest in using the Canadian Space Data Base.
28. Would you be willing to cooperate.
 29. What conditions.
 30. What benefits.
31. What type of organization structure would you favour.
32. What should be the federal role.
33. Is there a commercial opportunity.
34. Your general observations/questions.

APPENDIX 2
List of Interviewees

Interview List

- | | | |
|-----|---|--|
| 1. | Atomic Energy of Canada Limited
(INIS) | Mr. Mike Luke
Mr. Ozzie Tatone |
| 2. | CISIT Information Exchange Centre
(IEC) | Mrs. Maureen Gabe
Ms. Mary Simpson |
| 3. | CISTI CAN/OLE On-line
(OON) | Ms. Sheila Burvill
Mr. Michael Brandreth |
| 4. | Energy, Mines and Resources
(GEOSCAN) | Mr. Anthony Koph-Johnson |
| 5. | Canadian Transportation
Document System
(OOT) | Mr. H. Heyck |
| 6. | Micromedia Limited
(Microlog, CBCA, DAC) | Mr. Frank Gagne |
| 7. | Energy, Mines and Resources
(RESORS) | Ms. Marg Shurben |
| 8. | Sport Canada
(SPORT Database) | Mr. Gilles Chiasson |
| 9. | University of Guelph
(CODOC) | Ms. Virginia Gillham |
| 10. | Department of National Defence
(DSIS) | Mr. R. F. Nikkel
Ms. Janet Cathcart
Ms. Barbara Aubrey |
| 11. | Department of Supply and Services | Mr. Don Kelly |