EVOLUTION OF CANADIAN SYSTEM FOR GEOSCIENCE DATA, 1964-73 (EVOLUTION DU SYSTÈME CANADIEN DE DONNÉES GEOSCIENTIFIQUES, 1964-73)

Cornelius F. Burk, Jr.
Canadian Centre for Geoscience Data
Ottawa, Ontario K1A 0E8

ABSTRACT

A programme to apply modern information technology to the communication of geological data at the national level was begun in 1964. During the past nine years the concept of the Canadian System for Geoscience Data has evolved in response to advancing technology, and provides an illustration of how such technology can influence the conduct of science. In 1964 the Canadian System concept was predicated on standardized codes and formats for data; in 1967 standardized intellectual content was regarded as crucial; today, I consider systems intercommunication techniques as the key ingredient for a successful national system. Feedback to users from such intercommunication will promote a higher degree of standardization as a natural consequence. Thus, standards for data will become a product of information technology, not a prerequisite to using it. (Un programme visant à appliquer les techniques d'information les plus modernes à la communication de données géologiques au niveau national, fut mis à exécution en 1964. Durant les neuf dernières années, le concept du Système Canadien de Données Géoscientifiques a évolué devant le progrès de la technologie, et nous fournit une illustration de la façon dont cette technologie peut influencer la conduite de la science. En 1964, le concept du Système Canadien fut basé sur l'utilisation de codes et formats standards pour les données; en 1967, des "standards de contenu intellectuel" étaient considérés comme essentiels; aujourd'hui, je pense que les techniques d'intercommunication entre systèmes constituent l'élément-clé pour l'obtentien d'un système national efficace. Cette intercommunication aura pour effet, chez les usagers, de promouvoir un plus haut degré de standardisation. Alors, les standards pour les données seront le produit des techniques d'information, et non pas un prérequis à leur utilisation.)

GEOSCIENCE DATA

INTRODUCTION

The value of basic observations and measurements in the earth sciences is especially high in Canada, since in addition to playing their obvious role in the conduct of scientific research and development, geoscience data are among the *sine qua non* for development of our mineral and fuel resources and, increasingly, for the preservation of our environment. The current concern for energy sources, especially oil and gas, is but one example of an activity of deep economic and social concern in which geoscience data play a fundamental role.

In attempting to apply geoscience data more effectively to the solution of such broad problems, and to help advance the science in general, a small group of Canadian geologists began in 1964 to explore ways in which modern information technology could be applied on a national scale to the storage, retrieval and communication of geoscience data. Now, nine years later, it is instructive to review the developments that have transpired and to note, above all, how the availability of new technology has influenced proposals and activities related to the exchange of scientific data, and potentially to the conduct of science itself.

I wish to make it clear that I am using the much abused term "data" in the specific sense of basic scientific observations and measurements, that is, the records of what geoscientists or their instruments have seen or measured at some particular place and at some particular time. Such "data", with reduction and other processing, grade into higher levels of information including conclusions, syntheses and opinions. Although there is no objective way of distinctly dividing these various categories, the identification of the end-members in the context of a given study or investigation is usually clear. It is the "data" end-member of the information spectrum to which the Canadian System for Geoscience Data has addressed itself.

ORIGIN OF THE CONCEPT

At the 16th annual meeting of the National Advisory Committee on Research in the Geological Sciences, on 27 April 1964, K. R. Dawson of the Geological Survey of Canada described the contents and requirements of a proposed computer-based file for analytical and other data derived from laboratories of the Geological Survey. His presentation provoked a detailed discussion which included a proposal by A. D. Baillie, British American Oil Co. Ltd. (now Gulf Oil Canada Limited),

> "...that a subcommittee should be set up to make a survey of computer equipment and attempt to formulate a workable system of data compilation and retrieval that might be used not only by the Geological Survey, but by organizations across the country."

GEOSCIENCE DATA

This suggestion marked the beginning of a series of events which lead ultimately to the Canadian System for Geoscience Data. Thus, in late 1964 the concept of a "national system" and the feasibility of developing it were examined in a special study by S. C. Robinson of the Geological Survey of Canada. His unpublished report (Robinson, 1965) advocated the establishment of a national system based primarily on the use of 80-column punched cards for data storage and the establishment of a parallel national index to these data. Each data file would be controlled by its owner, but would include national standards for the data, including standardized reference numbers, a method of recording geographic location, and a standard classification and a method of description for rocks and other geological materials. Although third-generation computers were becoming available at this time, the rationale for these proposed standards was based on the capabilities of unit-record (card-sorting) machines, since this was the only type of equipment widely accessible to geologists.

CONCEPT DEFINITION

Following the feasibility study, the National Advisory Committee on Research in the Geological Sciences established an *ad hoc* committee under the chairmanship of S. C. Robinson to examine in detail the concept of a "national system" for storage and retrieval of geological data and to recommend on the means for implementation. A two-year study involving over 40 geologists and other scientists resulted in publication of interim (Robinson, 1966) and final (Brisbin and Ediger, 1967) reports which defined the National System and presented the results of pilot studies to illustrate how the concept might be applied in practice.

The concept which emerged envisaged an owner-controlled network of computer-based files, "linked" by the common use of various standards for geological data recorded therein. In addition, a computer-based national index to the sources of all data on Canada, computer-processable or otherwise, was included. Stress was placed on the decentralized, independant control of files in the network, and on the fundamental need for geologists to establish standards for their data. The proposed organization of the National System, as published in the 1967 report, is reproduced in Figure 1.

Use of the term "system" in this way was misleading to those who generally equate "system" with a tightly circumscribed set of procedures, usually operating under one roof. As presented in 1967, the National System could perhaps have been described more accurately as a "National Programme for development of compatible computer-based data files". Nevertheless, from a broader point of view, the concept did indeed constitute a system (cf. Canadian Postal System).

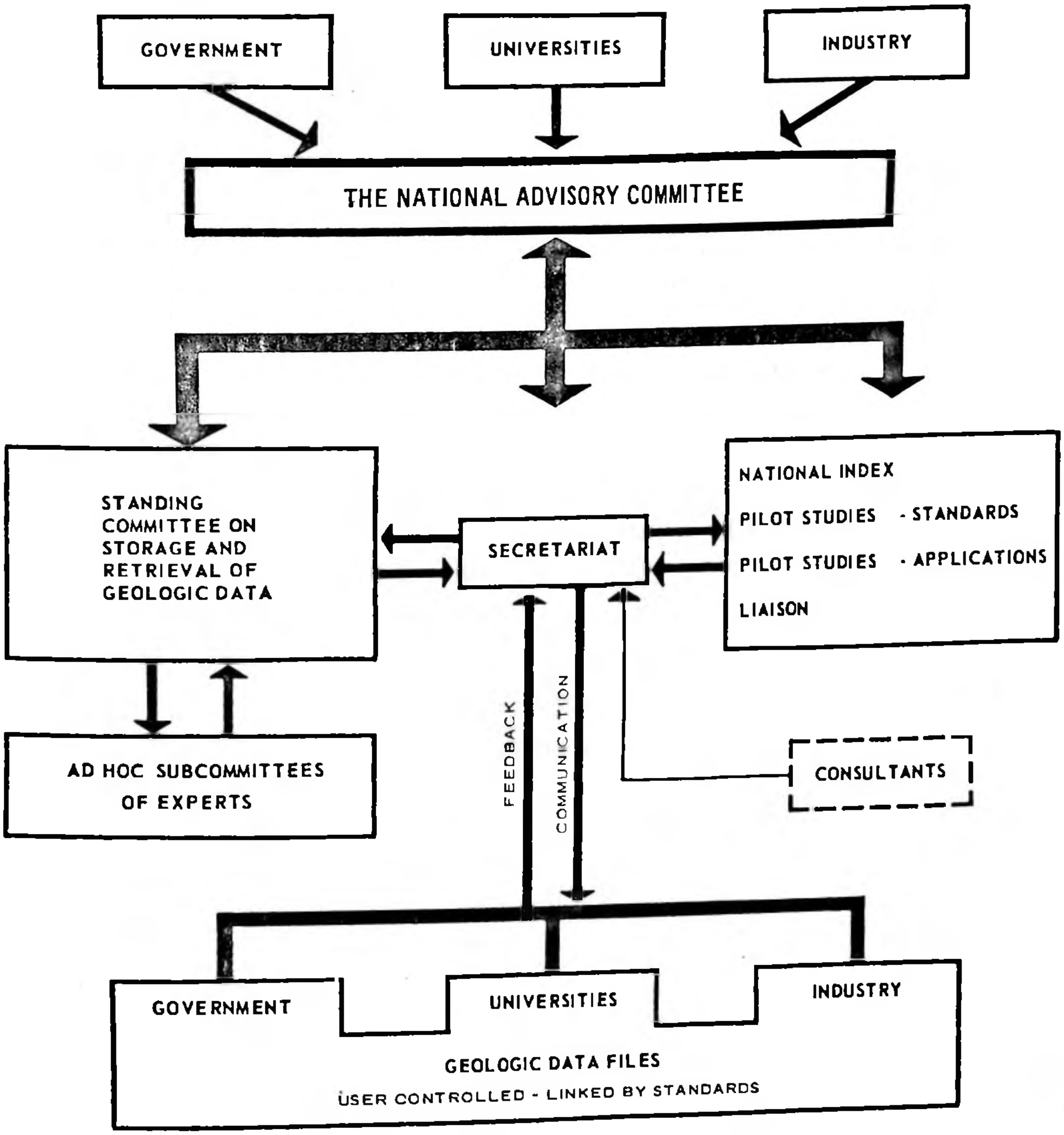

Figure 1. Organization of national system for geological data as proposed in 1967 by National Advisory Committee on Research in the Geological Sciences (reproduced from Brisbin and Ediger, 1967, Fig. 4).

GEOSCIENCE DATA

IMPLEMENTATION

Following discussion with various national bodies, especially the Provincial Ministers of Mines Conference and the Science Council of Canada, the Geological Survey of Canada began implementing the various recommendations of the National Advisory Committee's 1967 report for establishment of a National System. Among the major events were:

1. Permanent advisory committee established to oversee implementation i.e. NACRGS Subcommittee on Computer Applications (1967).

2. Pilot projects begun on computer-based national index (1967).

3. Special research grants awarded for developing computer-based data files in geology (1968).

4. Interim Secretariat for Geoscience Data established (1968).

5. Canadian Centre for Geoscience Data established to replace Secretariat (1970).

6. First edition of *Canadian Index to Geoscience Data* published (1970).

7. Advisory Board to Canadian Centre for Geoscience Data appointed by Department of Energy, Mines and Resources (1971).

During the period 1967-70, three major lines of activity were followed in pursuit of the National System concept, then renamed the Canadian System for Geoscience Data: Compilation of the *Canadian Index to Geoscience Data*, using systems software contributed by Imperial Oil Limited (SIS); development by committees of techniques and standards for building files of geological field data and mineral deposits data; and university research into data-base management and data files of various types. The development of standards for data was the primary goal behind most of this work, as had been recommended by the *ad hoc* committee. Although not overtly expressed at the time, the underlying reason for this goal - with emphasis on standards for data content, rather than codes and formats - was the assumption that available computer technology could not be used for the national communication of data on any significant scale without such standards. Of course there was also the desire to promote standardization on purely scientific grounds, but in the context of computer-based data systems it was felt that broad agreement on the intellectual content of geological files was basic to the operation of a decentralized Canadian System.

GEOSCIENCE DATA

CURRENT STATUS

The promotion and coordination of practical efforts to develop the concepts advocated by the National Advisory Committee report have been undertaken by the Canadian Centre for Geoscience Data since 1970 (Sharp, 1972). Its major activities during this period fall into three categories:

1. Inventory of data sources. Principally, continued development of the *Canadian Index to Geoscience Data*, with emphasis on various unpublished government reports (McGee, 1972), and more recently the identification and indexing of computer-based files (Gunn and White, in press; Hubaux, 1972).

2. Development of standards for geoscience data. Activity has been limited mainly to mineral and fuel deposits data, and to geological field data. For mineral deposits data, recommended standards are now being published (Kelly, 1972; Longe *et al.*, in press), but in the case of field data, emphasis has been placed on methods and techniques for developing successful computer-based systems for individual projects.

3. Coordination of data activities. The Canadian Centre for Geoscience Data has served as a focal point for inquiries, promoted contacts between workers, published bibliographies and other reports, and encouraged cooperation between government, universities and industry. The Centre also serves as secretariat for COGEODATA, the International Union of Geological Science's data committee. Research and development carried out under various auspicies, including its own, has been monitored and coordinated where possible.

Unfortunately, preoccupation with these day-to-day activities during the past three years has prevented the Centre from keeping in tune with an up-to-date conceptual framework. Technology has advanced so rapidly and in so many directions, that some of the basic assumptions underlying the Centre's programme of activities have become obsolete. The fact that there has been no general review of the Canadian System concept (Figure 1) since 1967 has, I believe, worked against the Centre's basic purpose of promoting the application of modern information technology in the most effective way possible. The Centre's Advisory Board will become more active in this area, but in the meantime I wish to present my assessment of the Canadian System concept as it exists today (Figure 2).

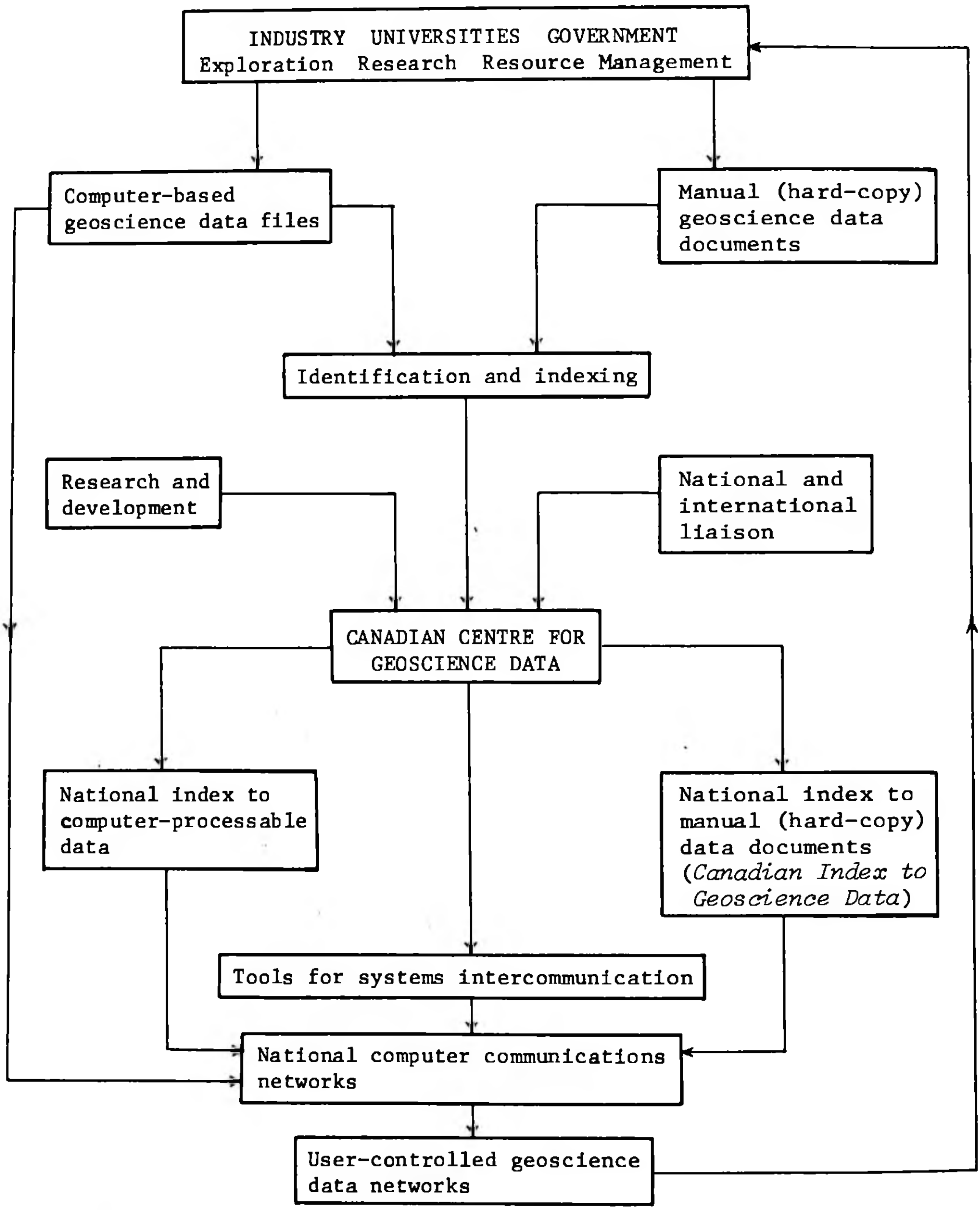

Figure 2. Present concept of Canadian System for Geoscience Data as seen by the writer.

According to this conceptual framework, which I emphasize is not necessarily "official" doctrine, the Centre is responsible for three main products. One of these, called "Tools for systems intercommunication", a phrase used by UNISIST (H. Brown, 1971, p. 41), refers to the wide range of methods, technologies, thesauri, guides, and to a certain extent, standards, by which two or more data files may be used for a common purpose. This function replaces the single element "standards", perceived by the 1967 *ad hoc* committee as the only key to developing a national network of independant files. During the past six years, however, we have witnessed the development of many new tools, such as generalized data management systems; computers have become even more powerful, and programming and systems skills have eliminated many elements of technical incompatibility which were at one time considered insurmountable. Increasingly, we are able to accept files "as they are", and nevertheless use them in concert. An outstanding example familiar to many is the CAN/SDI service of the National Science Library (J. E. Brown, 1972). Thus, there is an increasing tendancy towards the view, shared by this writer, that standards for geoscience data, at least on the scale suggested in 1967, are no longer a prerequisite for national network development.

The two other main products of the Centre are indexes to sources of data on Canadian geoscience (Figure 2). I have concluded that two separate indexes are required because of the differences in indexing approach required for hard-copy documents and computer-processable files, respectively. Moreover, the latter tend to be dynamic in content and form, thus requiring frequent review and re-indexing.

The blend of these basic ingredients - indexes to data and tools for systems intercommunication - when combined with privately controlled data files operating within our increasingly sophisticated national computer communications networks, will, I predict, result in achievement of the goal sought by the National Advisory Committee in 1964 - a national, owner-controlled network of computer-based files for the use of all who may require geoscience data.

EVOLUTION OF CONCEPTS

The notion that modern information technology could be applied to assist in the national communication of geoscience observations and measurements was first advanced in 1964. A succession of individuals, committees, and organizations has contributed to the development of concepts by which this could be achieved. The essential problem they have faced is: given the objective of a decentralized, owner-controlled national data network, which components of the total system require a national consensus and national application? The answers have evolved with time in response to advancing technology and have taken the following form:

Year	Key National Components	Technology	Reference
1965	Standard codes, formats, classifications, and descriptive terms, including geographic location. Standard reference numbering system.	80-column punched cards; card-sorting equipment.	Robinson (1965)
1967	Minimum standards for data content; standards for reference numbering, geographic location, and coding (if required). Permanent secretariat.	Computer systems; tape and disc storage.	Brisbin and Ediger (1967)
1973	File identification and documentation; tools for systems intercommunication (e.g. thesauri, tables of values, indexes, software); selected standards, esp. geographic location. Canadian Centre for Geoscience Data.	Computer systems; generalized data-base management systems; national computer communications networks.	This paper

The trend in this evolution has been away from the need for rigid standards for data content, notation and structure, and towards greater flexibility for data recording. Concomitantly, however, the same technology which is reducing the need for rigidity permits far greater practical use of data files among a wider audience. Such wide usage will, in turn, promote and bring about higher degrees of standardization among geoscientists. Thus, the application of information technology will, as a consequence, lead to greater standardization among geoscientists in response to legitimate scientific needs.

CONCLUSION

While it is generally appreciated that technology will continue to have a major effect on the evolution of national programmes for scientific and technical information, as illustrated by the Canadian System for Geoscience Data, the need for coordinating day-to-day activity with an up-to-date conceptual framework or model may be overlooked, and lead to serious delay and inefficiency. In practical terms, this danger can be avoided by the establishment of knowledgeable advisory and policy boards and by the maintenance of close working relationships between these bodies and practitioners of the art.

GEOSCIENCE DATA

REFERENCES

BRISBIN, W.C. and Ediger, N.M., *Editors*, 1967, "A National System for Storage and Retrieval of Geological Data in Canada", National Adv. Comm. Research Geol. Sciences (available from Geol. Survey Canada), 175 p.

BROWN, H., 1971, "UNISIST: Synopsis of the Feasibility Study on a World Science Information System", Unesco, Paris 1971, 92 p.

BROWN, J.E., 1972, "The Canadian National Scientific and Technical Information (STI) System; a Progress Report", National Fed. Sci. Abstracting Services, Rept. 4, May 1972, 26 p.

GUNN, K.L. and WHITE, J.P.W., *in press*, "MINDEX: An Index to Mineral and Fuel Deposits Data Files in Canada", Can. Centre Geoscience Data, Geol. Survey Canada Paper 73-16.

HUBAUX, A., 1972, "Geological Data Files: Survey of International Activity", CODATA Bull. 8, 30 p.

KELLY, A.M., 1972, "Recommended Standards for Recording the Location of Mineral Deposits", Can. Centre Geoscience Data, Geol. Survey Canada Paper 72-9, 8 p.

LONGE, R.V. *et al*, *in press*, "Recommended Standards for Data on Mineral and Fuel Deposits in Canada (tentative title)", Can. Centre Geoscience Data, Geol. Survey Canada Paper.

McGEE, B.A., 1972, "New Key to Mineral Information - Canadian Index to Geoscience Data", Can. Mining Jour., v. 93, no. 4, p. 43-47.

ROBINSON, S.C., 1965, "Interim Report on Possible Application of Data Processing Techniques to Storage and Retrieval of Geological Data in Canada", unpublished rept., Geol. Survey Canada, Ottawa, January 1965, 79 p.

________, 1966, "Interim Report of the Committee on Storage and Retrieval of Geological Data in Canada", National Adv. Comm. Research Geol. Sciences, Geol. Survey Canada Paper 66-43, 98 p.

SHARP, D.A., 1972, "The Canadian System for Geoscience Data", Proceedings 24th International Geol. Congress, sect. 16, p. 196-203.