SOCSCAN: A UNIQUE COMMUNICATION LINK AND INFORMATION SERVICE FOR THE SOCIAL SCIENCES IN CANADA

SOCSCAN: UN LIEN DE COMMUNICATION UNIQUE ET UN SERVICE D'INFORMATION POUR LES SCIENCES SOCIALES AU CANADA

Jake Knoppers
Director, SocScan
Social Science Federation of Canada
151 Slater
Ottawa, Ontario
K1P 5H3

ABSTRACT

The social science community in Canada has a widely disparate and diverse nature. At the same time, its component elements or disciplines overlap and intertwine but the linkages are linear or weak. This paper introduces to the information science conference the organization established to promote better communication and information linkages in the social science. In particular, this paper describes how a totally new concept in information retrieval, the integrated biographical and bibliographical data base SocScan running under a software package called ISIS has filled this communication and information gap.

RESUME

La communauté des sciences sociales au Canada est de nature très vaste et variée. De plus, ses composantes et ses disciplines se chevauchent et s'entremêlent, mais les liens qui les unissent sont à sens unique ou ténus. Cette étude présente à la Conférence des sciences de l'information l'organisation qu'on a établie pour faciliter une meilleure communication et de meilleurs réseaux d'information en sciences sociales. Cet exposé décrit, en particulier, comment un concept entièrement nouveau en recherche documentaire a comblé cette lacune dans la communication et l'information: il s'agit de SOSCAN, base intégrée de données biographiques et bibliographiques fonctionnant à partir d'un paquet-programme appelé ISIS.

SOCSCAN

INTRODUCTION

The Lamontagne Report on Science Policy stressed the need for rapid diffusion of new scientific developments and research-orientations especially since Canada spends very little as is on research and development (Lamontagne 1972). While Senator Lamontagne addressed himself primarily to what might be called the "hard" or "beta" sciences, the increasingly recognized reciprocal relations between the "alpha" and "beta" sciences in many areas such as environment, urban affairs, the Canadian North, etc., make his findings and recommendations more and more applicable to the social sciences each passing day. A later study by the Commission on Canadian Studies under the direction of Professor T.H.B. Symons emphasized the urgent need to develop the critical linkages between various fields of study relating to Canada (Symons 1975). Symon's recommendations apply equally to the whole area of the social sciences in Canada.

The social science community in Canada is an amorphous body. Lines of demarcation between the various disciplines are difficult to pin-point and one can even argue that they do not exist at all. For example, a chemical engineer fits in a certain niche or category but where does one place political sociology or anthropological linguistics. Further, while there are disciplinary journals and discipline-based associations or Learned Societies, the proliferation of interdisciplinary, subject, geographic area or time oriented studies, journals and associations during the past decade has tended to make categorization for information retrieval purposes already difficult even more difficult.

With reference to the theme of this conference "To better communicate information: a new step" and CAIS, the purpose of this paper is three-fold:

1. To acquaint CAIS with the existence and goals of the Social Science Federation of Canada.

2. To introduce the Canadian Directory Service of Social Scientists (SocScan).

3. To describe briefly how SocScan utilizes the software package called ISIS for information retrieval purposes.

SOCIAL SCIENCE FEDERATION OF CANADA

The Social Science Federation of Canada is a federation of academic social science associations and related organizations. It is a registered non-profit voluntary organization with the aim of promoting the social sciences in Canada. Specifically, the Federation's objectives are to:

- facilitate and initiate research in the social sciences, especially interdisciplinary research;

- contribute to the development of effective policies for social science research in Canada;

- develop the potential of the social sciences to contribute to the analysis and formation of social policies;

- represent effectively the social sciences in Canada and abroad;

- to develop Canadian initiatives and infrastructures for participation in the international context;

- administer programmes in support of these and similar objectives.

Specifically, the Federation consists of three categories of members. Ten academic social science associations or Learned Societies are the constituent members of the Federation. The number of associate members - universities, colleges, research centres and institutions and other non-profit organizations - presently 67, is growing. The third category are members-at-large. As many as 20 social scientists can be elected by the General Assembly for special contributions they can make to the achievement of the aims of the SSFC. All three categories are represented in the General Assembly which meets once a year. In conjunction with its annual meeting, workshops are held on policy issues and problems of special interest to the members. The SSFC receives an annual grant from the Canada Council (or now, the new Social Sciences and Humanities Research Council) as well as membership fees from its constituent members and contributions from its associate members (Masleck, 1977).

One of the programme established by the SSFC to assist it in meeting its objectives is the Canadian Directory Service of Social Scientists or SocScan.

CANADIAN DIRECTORY SERVICE OF SOCIAL SCIENTISTS (OR SOCSCAN)

In the early seventies, the Social Science Research Council of Canada, as the SSFC was then called, undertook to examine a proposal for the development of a Canadian Directory of Social Scientists. Such a proposal grew out of a central concern of the Council to respond to a crucial need to develop and disseminate adequate information on research activities in the social sciences in Canada and on human resources available for social science research.

The possibility of a printed directory had to be dismissed since the volume would be large with high printing costs. The long time lag between the reception of the information and the actual appearance of the printed directory meant that such a directory would be out of date before it even was printed. Further, at most 3 kinds of indices would be practical while updating would be awkward, expensive and infrequent. The printed directory approach, therefore, was too awkward, too slow in keeping up-to-date and too inflexible to meet all the needs of potential users.

The main concern in developing a directory service was that it would have to be an information retrieval system flexible and sophisticated enough to operate on a sufficiently detailed data base to meet the diverse needs of the social science community and the users of social science expertise. The decision was then made to look for a fast and adaptable computerized-data-base-search-approach. The retrieval system selected had to be able to perform a number of basic functions. Primarily, it had to be

able to select individual social scientists or construct groups of social scientists on the basis of any number of criteria. The major components of these criteria were to be comprehensive biographical data, language skills, areas of expertise or specializations, title and description of most recent or current research project (s) and a listing of the most recent publications.

Under the direction of Dr. Jan Loubser, Executive Director of the SSFC, development funding was secured, questionnaires were designed and market studies were undertaken in the fall of 1976 and early 1977. Concurrently, word processing software and information retrieval systems were being investigated and analyzed.

One major question that arose almost immediately was that of whether one should work in the environment of a controlled or an uncontrolled vocabulary. After an analysis of the nature and purpose of SocScan and the ways in which the end users would demand the product was completed, the decision was made to go the route of a free vocabulary. At present, there does not exist a recognized keywording or indexing system appropriate to the whole range of the social sciences covered by SocScan. Even if such a macro-thesaurus did exist at present, one problem which would immediately present itself would be whether to continue using keywording or indexing policies which no longer are of value to the end user which, due to changes in terminology and by their very a posteri nature cannot anticipate new research areas and interdisciplinary links.

Further, the data base contains both biographical and bibliographical information. For example, apart from most recent publications the respondents are asked to describe their areas of expertise or specialization and their current or most recent research topics. Because of the complexity, cost and time-lag in data-entry of information to which keywords would have to be attached to the ten major separate descriptive fields of eventually 10,000 social scientists, the decision was taken to go the route of what Svenonius called a "completely hospitable thesaurus" (Svenonius, 1976). (I am rewording her concept somewhat) Entry terms themselves would generate, by and large, the thesaurus or index terms the user can search on. All entry terms or words - after application of a stop-word list - are inverted for retrieval purposes. By using this technique virtually all words (except those which have been stopped) become access points in the inverted file. It would be interesting to do an analysis at some future date of the reduction ratio of entry terms to descriptor terms (or in our case more properly called, search terms).

The system chosen had to have the capability to accept raw data in the terms actually used by the researchers themselves. The data retained in the system should not need to be replaced or supplemented by general categories of code or keyword systems. One word processing and information retrieval system that can support such features is a software package called ISIS or Integrated Set of Information Systems.

SOCSCAN AND ISIS

According to Egyptian mythology, there existed at one time the goddess Isis who was the sister and wife of Osiris, the lord of the dead. Osiris was the son of Geb, the earth-god and together with Isis ruled over Egypt. Isis assisted Osiris in his beneficent activities for Osiris, also known as Horus when alive was the vegetation god. In the twenty-eighth year of his reign, Osiris was slain by his brother Seth and locked-up in a chest which was then thrown into the Nile. Isis seeking her husband in grief found the chest far away and brought back the body in a coffin. The myth goes on to tell how Seth hunting by moonlight, found the coffin containing the body of Osiris and cut the corps into pieces scattering them throughout the land of Egypt. Once more Isis renewed her search and recovered all the scattered members of her husband, except the membrile virile which had been devoured by the oxyrhynchus fish. Isis put the limbs together and with the assistance of her sister, the goddess Nepthys, performed magic ceremonies over the body and restored it to life. The risen Osiris, however did not remain on earth, but became the king of the "Western region", the place of departed spirits. (Hooke, 1971)

Thousands of years later, ISIS reappeared with the assistance of the International Labour Organization (Schafer and Brandeth, 1976). Osiris in the form of information, as vital to our modern society as the water of the Nile was to the ancient Egyptians, is still pursued by Seth. the dismemberer, who seeks to destroy complete descriptions. Faithful ISIS with the assistance of Nepthys, the search commands, restores and recomposes the scattered information segments which like OSIRIS eventually return to the place of departed spirits, the disk-packs and magnetic tapes of today.

Many and perhaps even most users approach and regard word processing and information retrieval services in the same way the ancient Egyptians regarded ISIS. The present-day ISIS supports a user-oriented on-line data entry processor. One enters the data in free format under the control of an interactive data entry processor after a field definition table has been established.

Search instructions are also phrased in a very simple language. The basic search is on a given field. The user may request all records containing a given word or phrase in the specified field. Successive search instructions may be combined in a variety of ways, either to expand or to contract the final list of retrieved records. Using its inverted files, the system can quickly locate the requested records. These can then be printed-out in Ottawa for mailing to the user. Any layout of the print-out may be specified by the user and the user may restrict the output to a selection of certain fields. (Schieber, 1972)

The nature and confidentiality of certain aspects of the SocScan data base preclude any direct on-line searches by the user. Further, a fixed set of predefined common keywords does not exist while often different meanings are assigned to the same terms or descriptors by the different disciplines. Therefore, requests for information are channelled and interpreted by the SocScan staff before being processed.

With the assistance and technical support of the International Development Research Centre (IDRC) and National Museums of Canada (NMC) it was possible to load and test a data base, now approaching 6,000 individual records consisting of an average of 6K characters each, within a short six-month period. (IDRC, 1972 and NMC 1977). Tests were run during the last two months of 1977 and SocScan became fully operational as a service on January 1, 1978.

Major directories of social scientists working in a specific area and indexes to their current research have already been produced on a per client basis on such topics as rural development, mental health, public administration, environment, energy resource utilization, drug and alcohol abuse, etc. Smaller more selective searches have also been done.

CONCLUSION

Even though additional records are being added, the SocScan service of the SSFC, is at present the most powerful and comprehensive tool available in Canada, and to the best of our knowledge in the world, for identifying and locating social scientists according to their discipline, area of expertise, language skills, qualification professional associations, current research projects most recent publications and so forth.

In a short two-year period SocScan has travelled from conceptualization to actual product permitting the social science community and related interests to have for the first time up-to-date information on itself and its activities. As such SocScan is the new national communication link and information service for the social sciences while at the same facilitating the analysis of trends in research, qualified manpower and mobility with in the social sciences and within disciplines as well.

REFERENCES

Masleck, Carolyn, Information Officer "The Social Science Federation" Fact Sheet, 1977 (mimeographed).

Fox, D.A. and Dink, R.A., "Statistical Tools for Data Base Study", paper presented at the 1977 CAIS Conference.

Hooke, S.H. Middle Eastern Mythology, Penguin, 1971 (first published 1963).

IDRC, Information Retrieval: ISIS, Ottawa, 1975

NMC, National Inventory Program, Ottawa, 1977.

Schafer, Marilyn E. and Michel Brandreth, "The International Development Research Centre and Computerized Retrieval Services", paper presented at the 1976 CAIS Conference.

Schieber, William D., Technical Manual on ISIS, Stockholm, 1972.

Svenonius, Glaine "Natural Language vs Controlled Vocabulary" paper presented at the 1976 CAIS Conference.