DATA ARCHIVES RETROSPECTIVE AND PERSPECTIVE*

N. K. Nijhawan

It is sometimes useful, in order to gain an understanding of the problems and opporunities for data archival efforts, to review the experiences of others. Especially the problems encountered in establishing archival activities in "Third World" nations may be illuminating for others, especially those used to the general level of technological deployment in the United States and Europe. N. K. Nijhawan has provided a description of current machine readable data archival efforts in India. Editor.

The object of this paper is to outline the programme of the Indian Council of Social Science Research for the development of social science data archives in India. It attempts to examine critically the work done so far in this regard and makes some tentative proposals for the developmènt of this programme during the course of the next few years. But before starting a deta¹led discussion of these issues, the meaning and purpose of data archives in general as well as factors that have led to the establishment and growth of such institutions all over the world during the past two decades or so will be briefly dealt with.

In the most general sense, a data archive is a library of data. Like libraries of books, which concentrate on the acquisition and cataloguing of books in order to make them accessible to the academic community, these institutions, known variously as data archives, data banks, data resources· libraries, etc., have been set up for the acquisition,
----

organization and dissemination of data to meet research and training needs of the social science community. Unlike usual libraries, however, the origin of these data archives is more recent. The formation of data archives started during the late 1950s and early 1960s when the computer became available to a sizable portion of the social science research community. These institutions have gained momentum during the past two decades or so, particularly in the USA and Western Europe, due to rapid developments in the field of computer technology on the one hand and use of quantitative techniques in social science research facilitated by this technology, on the other. Further, the urge to share data resources among social scientists has also contributed to the growth and development of these institutions. These developments, presumably, influenced setting up of this programme in India.

BACKGROUND

The V.K.R.V. Rao Committe which recommended the establishment of the Council also made a recommendation that one of the functions of the Indian Council of Social Science Research should be "to develop and support centres for maintenance and supply of data." In pursuance

of this recommendation, the Council set up a working group to suggest institutional arrangements for evolving an effective data archival programme in the country. The group submitted its report in March 1970 and made a number of recommendations, the most important of which concerned the setting up of a "data cell" in the secretariat of the ICSSR and the assisting of a few selected institutions in the country for developing a coordinated network of data archives. These recommendations were approved by the Council and the Data Archives was established in the office of the Council in New Delhi in 1973.

FUNCTIONS

The Indian Council of Social Science Research has taken an unconventional view of the functions of its Data Archives. Conventionally, as mentioned above a social science data archive is supposed to acquire, organize and disseminate machine-readable data sets for re-use. These are very important functions and were proposed to be performed by the Data Archives of the ICSSR. The Council was of the view that the Data Archives must assume some additional functions in view of the perculiarities of the Indian situation. For instance, we have a shortage of trained manpower. There is the problem of access to data produced by major data producing agencies. Equally important is the question of easy access of computing and adequate software facilities for social science research. Because of these problems, it was decided that the Data Archives should not only perform conventional functions of data acquisition and dissemination but also take up additional functions to help build a clientele

and get better returns from this infrastructure.

Briefly stated, the Council defined the functions of the Data Archives as follows:

1. to acquire, organize and maintain social science data sets in machine-readable form and make them available to interested social scientists for re-use;

2. to support suitable research institutions in different parts of the country to develop similar institution-based data archives;

3. to develop and maintain a computer programme library to handle the data management and retrieval problems of the data archives as well as data processing and analysis of problems of social scientists;

4. to organize training courses in survey research, and in data processing and analysis, with emphasis on the use of computers and other mechanical devices and computer programme packages;

5. to arrange for guidance and consultancy services for social scientists requiring assistance in data processing and analysis;

6. to act as liaison between official data producing agencies and social scientists in order to

incorporate the research interests of social scientists in data gathering and tabulation plans of these agencies: and

7. to establish collaborative relations with social science data archives abroad.

To this was added in 1976 the function of compiling a National Register of Social Scientists in India to be revised periodically.

DATA ACQUISITION AND DISSEMINATION

The question of types of data to be acquired by the Data Archives and the sources of these data were first taken up. Three main sources of data were identified: (a) data generated through the ICSSR-funded projects; (b) data generated by the official and semi-official agencies; and (c) data generated by scholars working in the research institutions and university departments. Again, keeping in line with the general objectives of the Council, it was decided to acquire both survey as well as aggregate data relevant to the research needs of all the disciplines falling under the rubric of social sciences. However, in view of limited resources, both financial and otherwise, the Council laid down certain priorities and gave the first priority to the acquisition of data sets generated by its own funded projects. The Council being one of the largest single social science research funding bodies in the country had at the time of the establishment of the Data Archives, funded more than 300 projects of which nearly 50 per cent had been completed. Since then, the number

of the ICSSR funded projects have more than doubled. It is, therefore, quite understandable that the Data Archives established by the Council would have a regular programme of acquiring and preserving data sets generated by its own funded projects.

During the past four to five years about 55 data sets have been acquired by the Data Archives. This also includes about ten data sets received from governmental agencies and other researchers who did not receive funds from the ICSSR. By any standard, this performance is not quite satisfactory though understandable.

It should be appreciated that the pace of developments, which appear to have resulted in the growth of such an activity elsewhere, took a late start in India and has been rather slow. The tradition of quantitative research, based particularly on secondary data analysis, is relatively new. A very small number of researchers make use of the facilities of computer technology and other mechanical devices for recording, processing and analysis of data. The availability of the hardware and software facilities for social science data processing are quite inadequate. What is worse, even the meagre facilities available are not being fully utilized because the community of social scientists is not oriented to and trained in the use of these facilities. Consequently, the majority of the scholars continue to analyse the data through hand tabulation and the raw data collected through the field survey are rarely transferred on punched cards. Therefore, only a limited number of useful machine-readable data sets are available for acquisition.

Generally speaking, factors mentioned above have been primarily responsible for the slow rate of acquisition to data sets. In the case of data generated by governmental agencies, however, lack of adequate resources, both financial and otherwise, have also contributed to the slow pace of data acquisition. For, over the years a large number of governmental agencies, such as, the National Sample Survey Organization, the Reserve Bank of India, and other governmental agencies have generated a vast magnitude of important data. The data are extremely relevant for social science research but not always available in a form in which they could be easily utilized by social scientists. It is, therefore, extremely difficult, if not impossible, for any single institution with limited resources to meaningfully organize these data and to make them available to interested scholars for re-use.

Consequently, the Council is gradually coming to the view that an all out frontal attack with close cooperation of the important data generating agencies in the country is absolutely essential to tackle this vexing problem. It is of the opinion that it is absolutely necessary to impress on all major official data producing agencies the urgent need to make the necessary arrangements for preserving their own data holdings, preferably in machine-readable form. Simultaneously, a group of social scientists as well as representatives of various organizations should work together and identify the nature, volume and the format of these data and help in evolving a phased programme for acquiring the most important data sets. These groups should also help incorporate the research interests of social scientists in the data

collection and tabulation plans of these agencies. No immediate solutions to these problems are in sight. The entire programme of acquisition and dissemination of useful data will depend on a number of factors, not necessarily within the control of the Council. Some of the activities which would, however, help in strengthening this programme, directly or not so directly will be discussed now.

INSTITUTION BASED DATA ARCHIVES

From the very beginning, the Council held the view that centralization of data archival activities in India is neither necessary nor feasible. For, a large number of research institutions and university departments, over the years, have generated important data. Most of these institutions do not have the necessary facilities for the preservation of these data. In the absence of proper storage facilities a vast amount of data is in the process of decay and death. In the majority of the cases, even information about these data is not systematically available to scholars outside a particular institution or beyond the personal contacts of an individual scholar. Appreciating these problems, the ICSSR decided to provide necessary funds and other services to selected institutions to help them develop data archives on a small scale.

The inadequate financial resources of the ICSSR have been the main handicap for the development of this programme. These difficulties are likely to continue for quite some time to come. Vast amount of financial resources would be needed to help these institutions purchase equipment needed for

the data organization and
dissemination. Additional funds,
on a regular basis would also be
necessaryfor carrying out other
oata archival functions. It may
not, therefore, be possible for the
ICSSR to provide adequate funds for
setting up full-fledged institution
based data archives. However, the
Council may consider providing at
least some funds to selecteo insti-
tutions having a large volume of
good data for upoating and organiz-
ing data and for preparing the
necessary documentation. Once this
is done, one set of these materials
may be made available to the ICSSR
Data Archives which will be made
responsible for the data dissemina-
tion task. Gradually, these insti-
tutions can be developed to take
over all the data archival func-
tions.

TRAINING COURSES

As stated earlier, the overall
success of the data archival pro-
gramme is closely linkeo with the
growth of quantitative techniques
in social science research in gen-
eral and use of mechanical devices
in data processing in particular.
These developments will generally
depend upon the overall system of
higher education in the country.
The structure of university educa-
tion will take some time to change
and respond to these needs. In the
meantime, the ICSSR is trying to
fill this void, to some extent, by
organizing training courses in
research 'methodology and survey
research techniques. Keeping in
line with these objectives, the
Data Archives decided to organize
training courses in the application
of computer technology and other
mechanical devices in social sci-
ence research with emphasis on use
of computer programme packages in
oata processing and analysis.

Young teachers and Ph.D. scholars
in social science disciplines are
invited to participate in these
courses of 3 to 4 weeks duration.
These courses have proved to be
quite useful and are expected to
continue.

GUIDANCE AND CONSULTANCY SERVICES

Quite closely associated with
the programme of training courses,
the Data Archives initiated in
1974-75 a programme for providing
guidance and consultancy services
to the social scientists to tackle
their problems in data recording,
processing and analysis with the
help of mechanical devices. In
order to make these services avail-
able to social scientists near
their normal places of work, a num-
ber of institutions have been
involved in this programme. Cur-
rently, these facilities are avail-
able, besides at the ICSSR Data
Archives through (1) the Indian
Institute of Management, Calcutta;
(2) the Centre for the Study of
Developing Societies, Delhi; (3)
the Sardar Patel Institute of Eco-
nomic and Social Research, Ahmeda-
bao; (4) the Centre for Development
Studies, Trivanorum; (5) the Gok-
hale Institute of Politics and Eco-
nomics, Poona; and (6) the Tata
Institute of Social Sciences, Bom-
bay. The Scheme is yet to gain
momentum at all the centres.

COLLABORATIVE RELATIONS WITH DATA ARCHIVES ABROAD

It would be readily admitted
that the data archival programme
cannot be aeveloped in isolation of
the developments in this field
abroad. It has been, therefore,
decided that the Data Archives

should maintain collaborative relations with similar data archives abroad for the exchange of information, data, software and even data archival staff. For various reasons, not much progress has been made in this field.

NATIONAL REGISTER OF SOCIAL SCIENTISTS

In the beginning of 1976, the Data Archives took up the task of compiling a National Register of Social Scientists in India. It was also decided that this Register should be revised periodically. This attempt is aimed at filling the void in the area of basic and comprehensive information about the background, research interests and contributions made by the social science community in India. Initially, it was proposed to include in this Register all the social scientists in university departments, colleges, research institutions, governmental organizations and private industry and business and to cover Anthropology, Commerce, Demography, Economics, Education, Geography, History, International Relations, Linguistics, Management, Political Science, Psychology, Public Administration, Sociology (including Criminology), Social Work, Communication (including Mass Communication and Journalism) and Law. However, because of practical difficulties the project had to be limited to the coverage of university departments, colleges and research institutions. The first volume of this Register comprising about 7,000 social scientists, information regarding whom was collected through a mailed questionnaire (covering the period up to December 1977) will be ready for publication soon. It is proposed to keep this information up to date, extend the coverage and

publish updated information periodically.

DATA INFORMATION SERVICES

All these functions discussed here are quite important and are proposed to be continued and further strengthened and expanded during the course of the next few years. However, these alone will not be enough. The scope of the Data Archives needs to be further broadened for providing better services. Systematic steps are necessary to build a data information system so that it can provide referral services in the sources of social science data in addition to the existing programme of physical acquisition and dissemination of important data sets. Making relevant, clean and properly documented data available to the interested scholar is an important task no doubt. And, no less important is it to provide information on the availability of relevant data to a scholar, even if the Data Archives might not be in a position to service that particular data to a researcher at a particular point of time.

In order to systematically build a meaningful data base for these services two programmes are proposed to be initiated. First, we plan to prepare an inventory of current and recently completed social science researches in India and keep this information up to date. Specifically, we intend to collect information on the status of current researches, type of data utilized, method of data collection, processing and analysis. Efforts will be made to cover all researches whether funded by the ICSSR or not. Second, we propose to initiate a series of projects

for preparing inventories on the type of data, periodicity of collection publication of data and unit of observation, etc., of social science data generated by official agencies. These two basic sources will throw open such a vast reservoir of information that when properly classified and organized, would prove tremendously useful in discharging this function. Moreover, this system would also help the data archives in identifying important data sets at the appropriate time and facilitate their acquisition.

DEVELOPMENT OF SOFTWARE FACILITIES

As has been emphasized earlier, all the data archives are supposed to acquire machine-readable data and to organize them in such a form as would make the data retrieval and dissemination least cumbersone and time consuming. In other words, data archives would generally acquire raw data transcribed either on punched cards, magnetic tapes or any other mechanical device. At the archives, the data normally have to pass certain 'acid tests', such as checks for wild codes, inconsistencies in coding punching, standardization of data formats and code categories, etc., before they become ready for dissemination. For all these tasks, a data archive needs relevant computer programmes (software). This software is used for data cleaning, transformation, organization and retrieval. This type of software is, therefore, necessary for the ICSSR Data Archives. In addition, the Data Archives has to develop and maintain another type of software if it proposes to cater to the statistical data processing needs of the social science community in the country.

Initially, the Data Archives had decided to develop this software on its own. However, not much coulo be achieved in this regard for various reasons. The most important reasons being non-availability of adequate funds and skilled manpower. Even if the funds were to be made available, the problem of a non-availability of properly trained programmers, who could meaningfully interact with social scientists, understand their requirements and develop the software accordingly, will continue to persist. Therefore, this programme will have to be developed over a long time. In the meanwhile, a beginning in this direction could be made by preparing an inventory of existing social science computer programmes in the country and making this information available to the interested scholars. Such an inventory would, decidedly, identify gaps in the availability of these facilities, both cross-sectionally and in different problem areas. Once such information beccmes available, the gaps that may exist, could be plugged in a phased programme under given priorities.

In sum, it may be recapitulated that the initial period has been a period of mixed experience. The achievements during this period, however, outweigh failures despite no prior experience in this field in the country. This period has been quite significant in many ways. During the period the necessary infrastucture has been built, some of the programmes put on the ground and sufficient experience gained to help take a leap forward during the course of the next few years.

However, it may be reiterated that the success of the data archival programme in the country would depend on the proper development of

various functions outlined above. Proper steps are necessary to strengthen these functions: conventional or unconventional, envisaged earlier or newly proposed. All this is important. And yet, the single most important factor for the development of this programme is an extensive use of quantitative techniques and automation in social science research. These developments are primarily going to come about through the changes in the university system of education in the country. These developments, so important for the growth of the Data Archives, are slow but sure to come about. In effect, it is in anticipation of these developments that the Council decided to initiate the data archival programme. However, in order to quicken the pace of development of this programme, the Council has initiated its programmes, like guidance and consultancy services and training courses in data processing and research methodology. The newly envisaged programmes, like the inventory of current research and software facilities would, hopefully, give additional support to this programme. Last, but not the least, it is the cooperation of the social science cummunity which will determine the rate and level of development of this programme in India.