Progress

Up to 1974, there appears to have been little interest in Australia in accessing secondary data. The Department of Political Science in the Research School of Social Sciences at the Australian National University held a Category C membership of ICPR from 1965 but access was largely limited to members of that department and certainly only to members of the university. The only data that was generally available at that time was the 1966 Census data distributed by the Australian Bureau of Statistics.

During 1974, four events occurred which provided the stimulus for a wider debate on the need for an Australian data archive:

i) a second department, the Department of Political Science at the University of Melbourne, became a member of ICPR;

ii) ICPR decided to reclassify Category C schools in Australia (and elsewhere) to Category B institutions for 1975-76 with a consequent increase in membership charges from $2000 to $3500 per year and further increases to follow. Almost simultaneously they proposed a new arrangement under which any number of Australian institutions could form a joint organization for Australia at a total annual subscription of $4000, subject to one of them acting as a clearing-house for the whole group;

iii) The Survey Research Centre was established at the Australian National University;

iv) Don DeBats, Senior Lecturer in American Studies and Politics at Flinders University, presented a paper to the Academy of Social Sciences recommending the establishment of an Australian data archive.
Without any one of these factors, it seems doubtful whether any progress would have been made for some time towards establishing an archive. DeBats had approached the National Library two years earlier with a suggestion that they take out a national membership of ICPR but his was a lone voice and it was felt that it would be hard to justify offering a new service, and one which would be a totally new departure in the type of material offered, when there was apparently no demand for it. At that time, the ANU was more concerned that any national membership should not adversely affect their own arrangements than with encouraging wider access. The proposed increase in charges and the presence of at least one other institution to share these charges and maintain them at the previously acceptable level provided the impetus needed for a national membership to be considered. The newly established Survey Research Centre had as one of its objectives the collection of information on survey data that could be made available for secondary analysis and was seen as the logical location for the national clearing-house.

As it was, following some preliminary investigation of possible alternatives and canvassing of the level of interest, a meeting was arranged for 16 February 1976 at the ANU and representatives of thirteen institutions attended. Eleven of these expressed an interest in joining an association of research and teaching institutions formed to take up a national membership of ICPSR (as ICPR was now called). This was taken out in May 1976 under the name of the Australian Consortium for Social and Political Research Incorporated (ACSPRI). Secondary objectives of this organization were

i) to collect and disseminate information relating to machine-readable social science data;

and ii) to investigate the desirability and feasibility of establishing an archive of Australian social science data in Australia or elsewhere and, if it is found desirable and feasible, to facilitate the establishment of such an archive.

Over the last six years, ACSPRI has grown from the two previous ICPSR members to nineteen member institutions at present, including twelve of the nineteen universities in Australia. Each member pays an initial joining fee of $150 and all share equally in the costs of ICPSR membership. Thirteen ACSPRI nominees have attended ICPSR Summer Training Programs. Data exchange agreements have been established with the Roper Center and the SSRC Survey Archive and agreement to redistribute data acquired from the Data and Program Library Service has been obtained. A Newsletter is produced twice yearly and distributed through representatives in member institutions and to other interested bodies.

Over the years, the number of orders for secondary data has remained stable at the level of about 9 a year, although the number of data sets distributed grew from 21 in 1976 to a peak of 85 in 1979 before dropping back to only 27 in 1980. In general, the pattern has
been that a member will place a large order for data sets soon after joining, and then orders will be for one or two data sets only.

Table 1. ACSPRI Membership and Level of Use

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of members at 31 Dec</th>
<th>No. of orders</th>
<th>No. of data sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1976</td>
<td>9</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>1977</td>
<td>10</td>
<td>8</td>
<td>46</td>
</tr>
<tr>
<td>1978</td>
<td>13</td>
<td>10</td>
<td>69</td>
</tr>
<tr>
<td>1979</td>
<td>16</td>
<td>7</td>
<td>85</td>
</tr>
<tr>
<td>1980</td>
<td>19</td>
<td>9</td>
<td>27</td>
</tr>
</tbody>
</table>

Although ACSPRI has been successful in providing Australian researchers with access to overseas data, it has been far less successful on its home ground. The ANU Survey Research Centre was relied on to undertake any data location and acquisition procedures but found that this was generally impossible due to its other commitments. As a result very few Australian data sets have been acquired to date. Excluding Australian census data, only three Australian data collections are available through ICPSR and only a further 24 data collections are available from ACSPRI.

If this situation had continued for much longer, I believe that membership of ACSPRI would have started to decline, probably quite rapidly. Already one of the founding members has dropped out because of lack of interest within the institution. For the great majority of academics, researchers or teachers, local data relating to local characteristics and issues is surely preferable to overseas data. In order to flourish, an archive must substitute for or add to the researchers' data collection activities, as well as provide new opportunities for data analysis, and these possibilities are more obvious with local data.

This situation now has a good chance of being rescued following the recent decision of the ANU to replace its Survey Research Centre with the Social Science Data Archives. The Archives will have a staff of six initially and should be fully operational early next year. In preparation for this, some preliminary investigations have been undertaken. In particular, sources of information on survey work in Australia have been examined and procedures to follow in acquiring, documenting, advertising and distributing data sets have been considered. The results of these deliberations and some of the questions they raised are presented below.
Locating Survey Data through Published Sources

1. Government Collections

"The Australian Bureau of Statistics is the official statistical organization for the Federal and State Governments. Its main function is to collect statistical information from a wide variety of social and economic areas and to compile statistics and disseminate them to interested users both within the Government and the community in general.

The ABS publishes currently almost 1900 statistical publications - either monthly, quarterly, half-yearly, annually or irregularly under approximately 700 different titles."

(ABS Catalogue of Publications).

The ABS is the major data collection agency in Australia. However to date, the Bureau has taken a very strict line on confidentiality of respondents and has been unwilling to release data in machine readable form in general and certainly not individual record data, de-identified of course. Data from the Australian Censuses of 1966, 1971 and 1976, with 1981 in a few years time, has been made available on magnetic tape aggregated at least to Census Collector's District level (an average size of 200 dwellings). From the 1976 Census in particular, Matrix Tapes containing counts of individuals or dwellings in cells of multidimensional tables were also made available, although in a format which required a considerable programming effort by the user to read and produce meaningful output. These data tapes are already held by the Archives. However, such important studies as the 1974 General Social Survey, the 1977-78 Australian Health Survey, the 1974-75 and 1975-76 Household Expenditure Surveys, the monthly Labour Force Surveys and many others are inaccessible and likely to remain so. Attitudes are changing however and there is some possibility of a sample of individual records from the 1981 Census being available for public use. In addition, the ABS will under certain conditions and when resources allow, conduct some analyses of individual record data on behalf of researchers.

Apart from the ABS, there are many other Government agencies at the Federal and State level who undertake data collection activities, and these agencies are generally more willing to make the data available to academic researchers. Until recently, information on these data collections was not widely available in any systematic form. However, Statistical Co-ordination bodies have recently been established by the Commonwealth and State Governments and each of these, with the exception of Tasmania, has compiled a register of statistical collections undertaken by the various Departments and Authorities of their respective governments. Entries are generally organized under the ABS Program Code or Department, and include the title, frequency, time period covered, availability and a contact officer. At the present
time each of these bodies uses a different data collection instrument and publishes its information in a different form, but there is some discussion of a unified approach for the future, provided that cuts in staff and available funds allow the continuation of these projects.

2. Opinion Polls

In the period 1941-1971, only one organization - Roy Morgan Research Centre Pty. Ltd. - conducted regular surveys of public opinion in Australia on an interstate basis. Two further polling organizations - Australian Nationwide Opinion Polls (ANOP) and Irving Saulwick and Associates - entered the field during 1971, and McNair Anderson Associates Pty. Ltd. began regular polling in 1973. The recent publication "Australian Opinion Polls 1941-1977" compiled by the University of Sydney's Sample Survey Centre provides a subject classification and keywords index to the questions included in the polls conducted by these four organizations up to 1977.

Data from about half of the 190 surveys conducted by the Roy Morgan Research Centre before 1968 are deposited with the Roper Center and can be made available to Australian researchers through a data exchange agreement between Roper and ACSPRI. Irving Saulwick and Associates' "Age Poll" is conducted in association with the Political Science Department at the University of Melbourne and permission has been given for these data to be made generally available two years after the completion of fieldwork. Negotiations are currently underway with the other three polling organizations to try to establish similar agreements.

3. Academic Collections

In a large and sparsely populated country like Australia it is very expensive to build and maintain a national fieldforce of interviewers for use in ad hoc surveys. As a result, any national surveys and the great majority of large regional surveys requiring personal interviews are contracted out to commercial market research agencies for the fieldwork. The only alternatives for large scale survey work are mail self-completion or other self-completion approaches such as surveys of school children conducted under supervision in the classroom. The vast majority of survey work conducted from the academic sector is however based on small samples from small geographic areas.

Information on the data collection activities undertaken by the academic sector is scattered through a whole range of publications such as annual reports of departments and institutions, reports of the granting bodies who provide funding for much of this research and the journals in which the results of the research appear. The need to provide some form of central register to these activities has been recognized in recent years and some progress has been made in this direction.
In 1975 the Social Welfare Commission produced the first edition of the Social Welfare Research Bulletin, which sought to provide a concise listing of Social Welfare research throughout Australia. Subsequently, the Department of Social Security took over production of this Bulletin and published updated versions in 1977 and 1981. Unfortunately, the latest edition is to be the last.

A number of other government departments provide bibliographic services on the areas of their particular interest. For example, the Department of Education maintains a Directory of Researchers and Research in Education; the Institute of Criminology scans publications for Australian or Australian-related criminological information and aims to collect copies of all publications relating to Australian criminology; the Department of Employment and Youth Affairs Library compiles quarterly bibliographies on a number of topics. However, the entries in these sources are generally limited to author, title and publication, and are thus rarely useful as information sources for the location of machine-readable data files.

The Survey Research Centre undertook two projects in an effort to provide more information on academic survey activity. The publication "Australian Social Surveys: Journal Extracts 1974-78" is based on a search of thirty Australian social science journals published in 1974-78 for articles reporting the use of survey data. Approximately 600 entries are organized under subject headings and include author, title, journal reference and, where available from the article, the geographical coverage, date, population, and sample of the survey. The second project, the "Inventory of Australian Surveys", was designed to provide more detailed information on survey work and used a mail questionnaire approach. Heads of social science departments in universities and colleges of advanced education were requested to give names and addresses of staff and postgraduate researchers who had conducted surveys from that department since 1970. Individual researchers were then contacted by mail and requested to give a detailed description of their work on an inventory questionnaire. Details of some 700 surveys are currently held on a computer file.

A comparison of the survey references attained in these two projects showed that both approaches suffer from undercoverage. Using details of the publications provided in the Inventory responses, a brief analysis of written items resulting from these surveys was carried out. Based on 617 entries, it was found that about one-third (210) of the surveys had not yet been reported at all, while about one-quarter (145) had resulted in journal articles. Of this latter group, at least 107 had published in Australian journals although only 69 were covered by the thirty journals selected for the Journal Extracts. To have located all of these references from a journal search would have required a doubling of the Australian journals covered, and inclusion of some overseas journals. Table 2 provides details of the types of written reports used.
Table 2. Written reports of surveys included in the Inventory.

<table>
<thead>
<tr>
<th>Written reports included in the Inventory</th>
<th>No. of surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>No written items reported</td>
<td>210</td>
</tr>
<tr>
<td>Journal articles - Australian journal</td>
<td>107</td>
</tr>
<tr>
<td>- Overseas journals only</td>
<td>19</td>
</tr>
<tr>
<td>- Others only - journals not checked, could be either</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>145</td>
</tr>
<tr>
<td>Books and monographs</td>
<td>63</td>
</tr>
<tr>
<td>Academic departments or institutional reports</td>
<td>83</td>
</tr>
<tr>
<td>Government and other reports</td>
<td>72</td>
</tr>
<tr>
<td>Published Conference proceedings</td>
<td>24</td>
</tr>
<tr>
<td>Unpublished Conference and seminar papers</td>
<td>27</td>
</tr>
<tr>
<td>Theses</td>
<td>95</td>
</tr>
</tbody>
</table>

N.B. Each type of written item reported counted for each survey.

On the grounds that there is clearly some time lag between the conduct of the survey and the appearance of a written report, an examination of written output by date of completion of fieldwork was made. Surveys resulting in theses, and those in which the dates of fieldwork or type of output was not specified were excluded. As expected, a higher proportion of surveys conducted before 1975 resulted in journal articles, but this was still only 41 percent of all these surveys, and 30 percent appear not to be written up at least 4 years later (Table 3). We have not as yet made any qualitative judgements about the merit of these surveys and it may of course be the case that it would not pay the archive to be too concerned about such work.

Table 3. Written output by date of completion of fieldwork.

<table>
<thead>
<tr>
<th>Completion of fieldwork</th>
<th>Written up in Journal Article</th>
<th>Written up in other form</th>
<th>Not written up</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 1975</td>
<td>41</td>
<td>29</td>
<td>30</td>
<td>132</td>
</tr>
<tr>
<td>1975-76</td>
<td>28</td>
<td>32</td>
<td>33</td>
<td>130</td>
</tr>
<tr>
<td>1977-78</td>
<td>21</td>
<td>29</td>
<td>50</td>
<td>184</td>
</tr>
<tr>
<td>After 1978</td>
<td>6</td>
<td>16</td>
<td>78</td>
<td>32</td>
</tr>
</tbody>
</table>
On the other hand, a journal search has some advantages in terms of coverage over the survey approach, due largely to the problems of non-response. From a sample of 103 survey reports in the Journal Extracts we found 9 with no address given and 13 in departments which had not been surveyed. Of the 81 remaining, 50 were not reported in the returns from departments, although 14 of these were conducted by researchers who were included in the Inventory for different studies. Of the 31 studies reported on the department returns, 22 summaries were returned by principal investigators.

Plans

While information on the data collected by government bodies, market researchers, academic researchers and other social science research bodies has improved considerably in Australia over recent years, there is a need to co-ordinate these activities and, if possible, establish a uniform approach. The concept of the Data Clearing House for the Social Sciences in Canada is I believe appropriate for Australia, although Canada has the advantage of a well established network of data archives. The Data Clearing House can thus concentrate all its resources on the provision of information services, for which (in 1975) it employed a full-time staff of six professionals, engaged in the developmental and service activities of the program.

The broad objectives of the Canadian Data Clearing House are:

1. the preparation of an index of quantitative social science data holdings that exist in machine-readable form and are to be found in Canadian universities, as well as in non-profit research agencies and other bodies conducting social science research;

2. the collection from federal and provincial government departments of a continuing description of their holdings and the performance of a liaison role between individual scholars and government departments;

3. the provision of information in response to individual inquiries, referring the inquirer to the source but not attempting to provide the inquirer with the actual data; and

4. the provision of technical information necessary for the more effective use of the data."


While our objectives have yet to be formulated and agreed, the provision of information about available data is surely necessary for determining a sensible acquisition policy. It would build on the work described above, although there is clearly a need to modify the information collection procedures used in our previous inventory work. A
balance must also be found between resources allocated to this activity and resources allocated to data acquisition, processing and dissemination, since we, unlike the Data Clearing House, will be attempting to provide the inquirer with the actual data. With this modification, the objectives stated above provide the basis for our planning at this time.

Inventory Plans

In recent years, there has been a strong movement among data archivists towards standardized documentation and increased bibliographic control of machine-readable data files (MRDF) with the hope that, ultimately, international union listings of available data may be produced. With this in mind, we felt that a new information system should be compatible with overseas developments where practicable. Although we have produced our own system for two bibliographies of Australian surveys, it is relatively unsophisticated and inexpensive to abandon at this stage. For all practical purposes, we are able to start from scratch.

In looking for a suitable description scheme, we required a form which could provide output in the form of a bibliographic citation; a title page; a full description of the study methodology, and content, and associated publications for inclusion in the codebook; and a more compact description for inclusion in a published inventory or catalogue of data holdings. Appropriate indices would also need to be generated by machine from the entry. The Study Description Scheme developed at the Danish Data Archives appears, with some reservations, to satisfy these needs.

The Study Description Form is essentially a more detailed version of the questionnaire used in our previous inventory work and consequently we are familiar with its style. The questionnaire used there was designed as an instrument which would be completed by the researcher, and returned to us for almost direct processing. In theory, our intervention would be minimal; in practice, it was not. To some extent it may have been due to faulty design, but returns required a significant amount of editing to give consistency, resulting in an untidy copy being sent for processing and thus more editing on the computer. It is therefore anticipated that the description for each study will be completed in-house, and will be based on published reports and other descriptive materials requested from the investigator.

Based on a very limited trial with three studies, we found only a few problems in completing the SD form in this way, although it is not entirely suitable for our purposes. Some sections in Part 2, Analysis Conditions, and Part 3, Reanalysis Conditions, will be omitted, and Part 5, Variables Included, will be compiled as listings of background variables and main variables/topics rather than use the categorized responses provided. The main reason for the latter is that we do not plan to implement a subject classification scheme immediately (preferring to wait for some recommended standard) and will use the main variables/topics as the basis for a keyboard index. As recommended
by users of the SD Scheme, Section 101 will be used to include the necessary elements of a bibliographic citation when these are not already included elsewhere, although an additional section has been added for details of the producer of the MRDF to provide a producer statement.

At this stage therefore it seems likely that the SD Scheme will be adopted by the Australian SSDA. There is however one reservation in our minds about adopting this scheme. At present, use of, and interest in using, the SD Scheme predominates in the European archives, with only one archive on the North American continent, the Leisure Studies Data Bank in Waterloo, using it. Clearly a standardized system has to be widely adopted to be a standard. Given the reported interest in establishing such a standard, we wonder whether the SD Scheme is being generally considered outside Europe, particularly in the United States; and if not, why not? Comments from conference participants on this topic will be very much appreciated.

Having chosen what we consider to be a suitable study description format, we are still faced with the problem of locating studies for inclusion in the inventory. As indicated by our previous experiences described above, providing reasonable coverage of the academic and other research agencies conducting social science research may be difficult. Again looking to the Data Clearing House model, a national network of designated correspondents and technical co-ordinators may be the answer and will certainly be tried. The SSRC Survey Archive also has a network of Archive Representatives covering university and polytechnic social science departments to publicize the Archive's services and acquisitions and to simplify request procedures. The basis for such a network is already established by the nineteen ACSPRI representatives, one for each member institution, and efforts will be made to expand and develop this network.

Compilation of the Inventory is seen to require three stages of information collection. Firstly, a record will be kept of current research and completed research comprising little more than names and addresses of principal investigators to be contacted, acquired through the network of representatives, reports of grant agencies and other information sources described earlier. Essentially, a mailing system for recording details of correspondence between the archive and investigators. Secondly, information on completed studies will be compiled from available publications and documentation supplied by the researcher. This will form the basic material, for deciding whether or not the data should be acquired. Prerequisites for inclusion of information at this stage is that the data is extant, in machine-readable form, and that the researcher is willing to make the data available to secondary users, perhaps conditionally, at some future date. Complete descriptions of data sets will only be made for studies acquired by the Archives and available from the Archives for secondary users.
Acquisition Plans

Acquisition policy will generally be determined by reference to the Users Advisory Committee which is being established for the Archives. Members of the Committee will be drawn largely from the social science departments of the University which include Demography, Sociology, Political Science, Economics, Economic History, Law, History, Urban Research and Statistics. In addition, at least one representative of ACSPRI will be on the Committee. Materials gathered in the course of compiling the Inventory will be presented to the Committee at regular, probably quarterly, meetings for a decision on the priority to be given to acquiring the data.

Highest priority will be given to acquisition in response to specific requests, which may be for a specified data set or sets, or for data relating to a specific topic. If the data are not already held by the Archive this will clearly involve some delay, but every effort will be made to minimize this. In the longer term, as the holdings of the Archive increase, the frequency of such requests should diminish.

The third basis for data acquisition relies on the attitudes of research funding agencies in Australia towards data archiving. The major funding bodies support a great deal of the primary data collection activity, particularly that of academic researchers, and should be supportive of an activity which will encourage wider use of these resources. Grant applications for additional funds to support the salaries and activities of additional staff will be made, which, if successful, will allow the Archive to develop more quickly. The Australian Research Grants Committee, the major source of academic research funding, agreed three years ago to include in its Advice to Applicants a request that social science data arising from funding projects be deposited with ACSPRI, but this has achieved little to date. Many overseas bodies make the deposit of such data a condition of grant, but this has so far been resisted by the ARGC. The Department of Health has this year provided funding to support the establishment of an archive of survey data on drug use in Australia and this project is underway.

There are I believe major advantages in focusing data acquisition on specific substantive areas where funding is largely centered on a single agency. The problems of locating suitable data can be overcome through reference to the agency's records, and the agency's involvement may act as an inducement to researchers to deposit their data. A substantial collection of related data provides greater opportunity for secondary analysis, and the agency supporting the creation of an archive will surely want to encourage use of the resource which in turn would encourage further support of the archive from the agency.

The Archives' Users Advisory Committee will also decide the level of data cleaning to be carried out on data acquired. On receipt of the data by the Archive, a minimum level of range checking will be done, and where necessary, multi-punch data converted to single-punch. More detailed checking of the data, error corrections, and creation of a codebook
by the Archive will only be undertaken on studies thought to warrant the effort and expense.

On a general point of inter-archival co-operation, it would surely benefit all archives, and new archives in particular, to have information readily available on the types of data set most often requested. The SSRC Survey Archive provided us with a list of their 25 most heavily demanded data sets which they concluded "demonstrates that national and cross-national rather than local surveys, and longitudinal panel and time series rather than one-off surveys, attract the heaviest use." I feel sure we would all like to know whether this is a general conclusion or one which is perhaps a result of the particular holdings of the archive at the time. British Election Studies and Family Expenditure Surveys form a significant part of their list, but does this reflect the substantive topics of interest or the quality of the survey work or some other factor? Many established archives will surely have conducted user surveys and it is important that the result of these surveys be widely available to all archives.

Dissemination Plans

To date, formal advertising of ACSPRI services has been done through distribution of the ACSPRI Newsletter. Editions of the Newsletter are produced in March and September and distributed by the ACSPRI Representative largely within their own member institutions. My intention in establishing the Newsletter was to carry reports of research and teaching applications of secondary data from contributors, but unfortunately no such contributions have been received over the two years of publication.

ICPSR provides ACSPRI with seven copies of codebooks for all Class I data sets and these are distributed to codebook centers located around Australia, one to each state. Each ACSPRI Representative receives a copy of the ICPSR Guide to Resources and Services and Information Mailings, and researchers wishing to consult codebooks can borrow them from the nearest codebook center. Of course this places researchers at any but the seven institutions with a codebook collection at some disadvantage, but the cost of establishing more of these centers would be considerable. Seven points of access to the codebooks is nevertheless clearly preferable to only one.

With the establishment of the Social Science Data Archives, the primary task will be to provide information on and access to Australian data as opposed to data from overseas archives, and to broaden the interest in secondary use of this data. As reported above, attempts will be made to extend the network of representatives down to departmental level as opposed to the current institutional level, and to include more institutions in the network.

The principal output from the archive will be derived from the study summaries compiled for the Inventory, since it will contain details
of many more studies than the archive has in its holdings. For studies which have not been acquired, entries will exclude specific details of the principal investigator to avoid the possibility of unsolicited direct approaches. Copies of the Inventory will be distributed free of charge to department representatives and be made available to libraries and individual researchers on a subscription basis. For studies held by the archive, documentation will be distributed to ACSPRI member institutions free of charge, but otherwise sold at cost. The Newsletter will continue as the main publicity medium, being distributed free through the local representatives. Data requests will be charged on a fee-for-service basis.

Summary

There are a number of alternate ways to establish and develop a data archive and we are faced with choosing one of them. Essentially I see a data archive as a consumer-oriented marketing activity with the academic social scientist as the primary consumer, the archivist as the marketing manager and data sets as the primary product. The product is not manufactured by the archive but is picked up second-hand from other sources. The archivist has the job of locating suitable products and deciding which to acquire, and whether or not it is worth cleaning up what is acquired before making it available to the consumer. The problems facing our marketing manager are:

- what data sets to acquire and in what quantities?
- where to acquire the data sets?
- which data sets should be cleaned?
- what promotion activities should be undertaken?

with the object of maximizing the consumer awareness and use of the product subject to the constraints of the limited resources available.

The marketing manager realizes of course the need for information on which to base these decisions and, being the manager, delegates responsibility to his market researcher. She (in this case) carries out a literature search and, since this is a new product on the Australian market, contacts similar marketing operations overseas requesting relevant information. Unfortunately, neither source proves very fruitful.

The marketing manager is thus placed in something of a dilemma, and decides to take a cautious attitude. There seems little point in filling the warehouse with materials which may never be sold - this would simply be doing something for the sake of it. On the other hand, it may be that by filling a warehouse with goods, chosen because they are readily available, and having a good advertising campaign, enough interest could be generated to clear a lot of it even if it was junk for the most part. On balance though, he feels that the consumer market he wants to attract is fairly discerning and that, although they may initially be attracted to the warehouse, their disappointment with the available product will discourage any future interest.
Taking this view, the manager decides that first priority should be given to establishing a good network of contacts among the producers, creating an information source on the availability of goods of interest. The producers themselves are of course interested in the activities of fellow producers and it is felt that their co-operation would be gained by offering them the results of the information collection in exchange for their involvement, in the way that estate agents pool information on houses for sale in multi-list schemes. The producers here are also the most likely consumers and the information system will both assist them in planning any new product and encourage their interest in the products of others.

Acquisition during this initial phase will not be substantial, being concentrated on satisfying customer orders, which are also unlikely to be substantial, and pieces of particular merit selected by a board of expert advisors. These special pieces will be used as the center-piece in promotion activities, and seminars and workshops will be devised around them.

In the longer term, obtaining input to the information system should become less demanding of the archive's staff allowing redeployment of resources to promotion, cleaning, new acquisitions and distribution activities. With what is essentially a new product on the Australian academic market, promotion must be given high priority in order to attract new customers and to keep old customers up to date with new products. Information gained from the network of producers and consumers and orders placed during the initial phase will provide a guide to customer requirements, allowing effective planning of and control over future acquisition and cleaning activities.

LABOR STATISTICS FOR SALE ON TAPE

NTIS, the National Technical Information Service, has available on magnetic tape statistics from the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor. The LABSTAT database includes:
1) manpower information such as labor force characteristics, employment hours and earnings, nationally and by SMSA, unemployment data by SMSA and labor turnover, 2) the Consumer Price Index, Producer Price Index, and Export & Import Price Indexes, and 3) imports statistics, value by industry. Each series is updated monthly and is available as a demand item or by subscription. For pricing and ordering information contact:

Stuart Weisman
Product Manager
(703) 487-4807
Societies everywhere are being affected by the new information technologies. In addition, they have become increasingly dependent on statistical information for making important public policy decisions. Social science data services, a direct result of the new technologies, have been established to provide easier access to computerized statistical information. One would expect therefore to have seen over the last 15 years a great many data services established throughout institutions of higher learning and government. Yet these data services are few and the ones that exist, underutilized. There are obviously many reasons for their underutilization.

Today, I will address three reasons which contribute to the current situation. Poor quality data impede good decision making and research. Lack of coordination and planning of the statistical information system make it very difficult to produce, locate and retrieve data. New information technologies are modifying our societies. But social scientists are not directing enough attention to how society is being altered and we lack appropriate models and data. My concluding remarks suggest a number of ways that social scientists can contribute to improving the current situation.

Social science data archives and services, like their predecessor libraries and archives of print documents and film, represent one component of a society’s institutional memory. The underlying philosophy of preservation and access holds that transfer of the data collections from their producers to these data centers greatly increases the return on the original public and private investment.

This paper was delivered at the 1981 IFDO/IASSIST conference in Grenoble. The author gratefully acknowledges helpful comments on earlier versions from Thomas Flory, Nancy McManus, and Richard C. Roistacher.
Most of these centers were established before their national archives created machine-readable divisions. Although these centers have not been designated official repositories for government records, governments have turned to them for assistance in retrieving government data files. As recent experiences in several countries demonstrate, more government data producers are delegating archival responsibilities to university data repositories in recognition that government cannot preserve and maintain its own records.

As society's problems have grown more complex, statistics have become more important to effective decisionmaking. Not only do policymakers face increasingly complex issues, but many problems now interact with one another (12,136).

The resources of data centers, for holding historical collections of data and for generating new ones, are essential if national policy decisions are to be made in a more rational manner. Existing administrative records systems, used for secondary analysis or linked to new data collection activities, provide a means for responding efficiently to new policy questions.

Data services are also important for cumulative social science research activities. Common access creates a "commonality of research among widely separated scholars" (9,411). The data archive acts as a scientific laboratory which encourages the sharing of data, multidisciplinary exploitation of evidence, and "multiple and complex analytic applications" (5,393). The data center makes a pedagogical contribution by allowing the student to participate in scientific inquiry, developing problem-solving techniques and behavior like those of students in the natural sciences. A recently completed study of factors influencing the sharing of computer-based resources for higher education and research shows a direct connection between utilization and sharing. It suggests that the "seemingly indirect attempts to broaden 'computer literacy' and computer use might have systemic effects on the level and nature of computer-based sharing" (8,4.44).

Less obviously, the data archive plays a role as an agent for assessment of information transfer activities. It offers administrators and researchers the opportunity to assess the technical, administrative, economic, and policy issues related to standards of data quality, documentation, access, and distribution.

Nevertheless, 33 years after the Roper Center at Williams College in Massachusetts and 20 years after the establishment of the Steinmetz Archives in Amsterdam, the Zentralarchiv für Empirische Sozialforschung at the University of Cologne, and the Inter-university Consortium for Political and Social Research at the University of Michigan, no more than 50-odd data services exist throughout the world, almost all university based. National governments have been slow to accept the idea that data services play an important role in information policy development.
Information technologies and services produced and offered by the private-for-profit sector are beginning to dominate access channels.

Why are there now so few social science data archives and why do they appear to be underutilized? That they have been is due to a wide array of reasons. Rather than providing an inventory of these reasons, I will address the complex and interdependent issues of the quality of statistical data, factors responsible for the lack of coordination of data resources, and the need to make social science more relevant to policy choices.

Most of my remarks have been stated in one form or another during the last five years in many countries. I address the creation of statistical data and administrative records produced by government because it is a major provider of the data resources which social scientists use. And I expect that in the future, government's influence on statistical data production will determine even more how the social scientific community conducts itself.

My remarks about the role of social science in an Information Age have been influenced by recent political events, in which many questions have been raised about the relevance of social science. I believe that relevance implies and requires philosophical reflection. Relevance requires use of theoretical perspectives about human and social interests. Relevance requires new models which integrate our natural and social worlds with scientific and technological discoveries. My recommendations for improving data quality, planning, and coordination should be understood as two aspects of the larger philosophical and moral dilemmas which we confront. Thus, the last part of my address reflects on some of the questions social scientists must seek to answer as they confront social changes which are the result of new information technologies.

II. Problems of Data Quality and of Coordination and Planning

A. Data Quality

Dissatisfaction with the quality of data is widespread throughout the scientific community and government, although enormous strides have been made to improve measurement. David R. Lidd, Jr., director of the Office of Standard Reference Data at the U.S. National Bureau of Standards, recently wrote "that a considerable amount of information in such archives is erroneous". He cited almost 200 reported measurements of the heat conductivity of copper—"a range of values so great that most of the data are clearly off the mark" (12). Publications of social and science indicators, on which many projections in the United States are based, contain obvious statistical errors—obvious, that is, once the data are examined—and inadequate information on sectors of the society which we know are undergoing rapid changes. These errors are due in part to inadequate sampling frames and improper methodological tools applied to data gathering and analysis.
Three factors that influence the quality of statistical and other data and their analytic potential are demand (or user requirements), supply (or the resources of the system), and structural or environmental conditions.

User Requirements. A recently published White Paper on the U.S. statistical system notes that "the complexity and urgency of issues facing policymakers often leads them to demand more data and more timely data, with little regard for quality" (11,164). Policymakers tend to be uncritical about the quality of the data they use; social scientists only somewhat less so. The immediate demands for completing the administrative function, a budgetary horizon of one to two years, and legislative demands for information for modifying policy impede the necessary gestation period for designing and gathering data. Political ends influence the quality of data. "... Some of the most important statistics are held hostage to political ends by their visible and direct use in politically important decisions which allocate (national) resources" (4,204).

Resources for Maintenance and Improvement of Quality. At least in the U.S., there has been no thorough government-wide review of classification standards for statisticians for about three decades. Professional training in data handling is received (or not received, as the case may be) on the job, with little influence by non-governmental sources of expertise. The social science community, which has discovered many useful tools for improving data quality, has little opportunity for interaction with the governmental data producer and statistician. This interaction is not encouraged by government and the university organization nor by attitudes of the government administrator or academician. Civil servants' opportunities for career development and participation in conferences such as this one are limited.

The White Paper offers other explanations. Budgets for statistical programs and projects do not include resources for internal and/or external measurement of quality. Funds are seldom provided for methodological research to improve quality, except where there are clear indications of serious deficiencies. Such deficiencies may not become obvious until the effects of poor policy decisions are felt. Political bodies are then moved to apply remedies (which rarely reflect the underlying systemic problem). Little attention is given to the basic design of surveys, evaluation studies, program experiments, and data bases developed for policy analysis. Competitive procurement activities (contracts, for example) seldom receive adequate technical review, and selection panels often lack the technical skills to make an informed judgment (11).

A 1978 study by the U.S. General Accounting Office of federally-sponsored attitude and opinion surveys found serious technical flaws which limited the usefulness of the results in all five surveys which were reviewed in detail. The GAO concluded that "better guidance and controls were needed to improve Federal surveys of attitudes and opinions" (11,162). Another study, sponsored by the American Statistical Association
and funded by the U.S. National Science Foundation, evaluated 26 sample surveys conducted in 1975 and found that 15 of the 26 surveys had serious technical flaws. All but two of the 26 federally sponsored surveys were conducted under contract by universities or other private survey research organizations (1).

Structural factors affecting quality. Increasingly, statistical services are being procured from outside the government under contract. Agencies often have funds to acquire these statistical services, but no budget to develop staff and inhouse organs to build services and decide on technical specifications and selection. Operations which include data collection by other units of government are notoriously difficult to monitor and to standardize. For example, a large portion of the data collection activities conducted under the auspices of the intergovernmental Cooperative Health Statistics System program in the U.S. is being eliminated; quality control was cited as a major factor in this decision (13). Producers outside government are typically unaware of the uses to which their data will be put, or of the utility of the data they provide or of the administrative needs of an agency. Analysts are often unaware of important limitations of data because technical standards of data description have not been instituted by government agencies.

Restrictions on interagency sharing often result in the lack of comparability in data produced by different agencies. Such restrictions sometimes result in failure to fully exploit expensive data bases. Although policy may require linkages of materials gathered in several agencies and from several records series, legal procedural, and operational mechanisms to provide linkage are few and far between(2).

B. Lack of Coordination and Planning

Poor information management practices applied to statistical and administrative records and the internal organization of bureaucracy are in part responsible for difficulties in accessing records. These problems have led "to a growing incidence of overlap, duplication, mismatch and gaps in data and analysis, and increasingly complex problems of access by users and statistical agencies to various Federal data" (11,143).

Nora and Minc give three examples of this kind of compartmentalized development in France.

Hospitals have developed systems for billing medical expenditures and hospital-stay expenditures without collaborating with Social Security. Within Social Security itself, compartmentalization into three branches, each with its own data processing centers, has led to manual retrieval of data produced by the computers of the other branches. As a result of the present departmental separation (they write before various reorganizations within the Mitterand government), the Direction Générale
des Impôts and the Direction de l'Amenagement Foncier et de l'Urbanisme (land development and urban affairs) has each established a land use data bank, the former for tax purposes, the latter for development purposes. The legal definitions and the types of information differ. Nevertheless, there are broad common areas, but nobody worries about them. In addition to the waste, the establishment of these two data banks prolongs administrative isolation. Strengthened by this investment, both administrations are prepared to resist attempts at rapprochement (10,115).

Within the U.S. government, the Federal Trade Commission in its quarterly financial reports asks for data which are available in quarterly filings with the Securities and Exchange Commission. And there are currently three duplicate mortgagee interest surveys (11,149). In Wisconsin, the Department of Public Instruction refuses to turn over computerized records that the Department of Revenue needs for statistical analyses and modeling. The Department of Revenue is forced to collect this information manually if it is to perform its work in a timely way.

The application of data processing technologies has been uneven throughout government, and as Nora and Minc note, although "penetration has been extremely rapid," it has "taken place in uneven ways, strengthening barriers, immobilizing the structures that it penetrates for a long time" (10,112). They note that

in the majority of cases, each department acquires data processing capabilities without worrying about the possible difficulties that its plan may cause elsewhere, and especially without measuring the "synergistic" effects that better coordination with other departments might have produced (10,112).

The high rate of change in administrative data processing has resulted in a phenomenon that could be called input without throughput. Delays in the implementation of data base management systems, complications in electronic data entry systems, pressures to maintain routine administration in the face of high staff turnover in data processing, and the imposition of computer technology on organizations designed for manual systems have created serious bottlenecks in routine administration. Procurement policies emphasize centralization and are costly and a serious impediment to acquiring the most economical and efficient technology available. Little attention is given to identifying areas where decentralization of the information system would improve an agency's capabilities. On the other hand, administrators have few possibilities and little incentive to improve coordination because statutes delimit an agency's mission.

Even when research access to identifiable information is not in question, attention has not been given to maintenance and preservation of machine-readable records. Constraints on administrative activity
tend to reduce incentives for "backward" looks, those that would require that records be maintained and preserved. The resulting costs can be very high. For example, efforts now underway to create public use samples from microfilmed versions of the 1940 and 1950 U.S. Censuses of Population are to cost $8 million. Much of that information was on punch cards at one time. Records managers and archivists do not usually participate in decisions about retaining and destroying computerized records. As a result, computerized records are not integrated into records management practices.

Records managers leave decisions about retention to those with programmatic responsibility and concern themselves with managing paper and microfilm records. Records and computer centers see themselves as repositories for magnetic tape, with responsibility for decisions about tape maintenance left in the hands of an agency. Individual analysts retain information on the contents of files for which they have programmatic responsibility. Data processors are often the only persons knowledgeable as to format and physical attributes of computerized records. Documentation for MRR may not exist or may be scattered among the various agency personnel responsible for the different aspects of MRR. Valuable data are routinely erased and the tapes are reused when tape shortages occur, often without prior systematic review.

III. Society and the New Information Technologies

The emerging information technologies are already altering the nature of our society and affecting existing political, economic, and social institutions and values. Data processing is accelerating production,

with less but more effective work and jobs very different from those imposed by industrial life. This change has already begun: a great decrease in the labor force in the primary and secondary sectors, an increase in the services, and above all, a multiplication of activities in which information is the raw material (10,126).

Already, computerization of formerly manually performed tasks is rendering the semi-skilled and unskilled worker unemployable. Robots are beginning to replace humans, performing certain tasks more efficiently and increasing industrial productivity. However, not only the unskilled or low-skilled are being replaced. The introduction of automation is affecting highly skilled technical workers. For example, although more than 12,000 air traffic controllers walked off their jobs in the United States, air traffic was only partially reduced because computers assisted in air traffic decision making. In the opinion of some, computers were used as a strike-breaking tool(3). The Federal Aviation Administration hopes within 10 years to have computerized en route air control to such an extent that at least 50% fewer controllers will be needed and those that will be needed will be computer managers(6).
Economic changes will be accompanied by a change in the structure of organizations and by fluctuations in attitudes toward work. As numerous examples have demonstrated, the new technologies related to automation and data processing can flourish in small as well as large organizations. The psychological and social bonds that were created by the work place and that fostered worker solidarity will weaken as automation enforces isolation.

Monetary and other rewards will go increasingly to those who have the means to produce and manipulate the technology, creating new elite structures and placing political decision-making in the hands of technicians. As Duncan McRae has noted, the "risk of technocracy lies in the possibility of uncontrolled power held by an elite and devoted to special values and interests rather than to the general welfare" (7,45-46).

IV. Recommendations

In what ways can social scientists contribute to improving the present environment of the information system? The information system in which statistical data production and analysis take place is highly complex and dependent on new technologies. It requires expertise from many disciplines and specializations. It requires modifications in the institutional framework in order to cope effectively with societal change and to anticipate unexpected policy and political demands.

The social scientist and policymaker have many common interests. They have a great deal to gain by cooperating, to improve the quality of data, coordination and planning, and access to computerized records. Governments must use available expertise "in data collection and analysis activities, starting at the design stage, and continuing through to evaluation of how results are used" (11,166). Social scientists can contribute through methodological research in measurement of errors to improving collection methods and to improving the presentation of information about methodology structure and other limitations of the data products and analyses. The results of methodological research must then be widely disseminated so that they can be evaluated, criticized, and competing methods proposed if necessary.

We must be concerned with creating an integrated output and with producing cross-cutting analyses over a wide range of issues. Social scientists can assist in substantive integration activities, by developing standard concepts, definitions, classifications, survey frames, and procedures, and by monitoring and promoting their utilization by government and by the private sector. Social scientists can assist in developing a "consistent conceptual framework or model based on behavioral relationships in various disciplines" (11,172).

There needs to be increased use of administrative records to produce statistics and to respond to public policy questions. Public use samples should be drawn from administrative records. Administrators should be
made to produce public use files and to coordinate record linkage and analyses. Through their activities, social scientists can promote record linking at the microlevel and demonstrate ways in which the data's analytic potential can be enhanced. (It is important to note, by way of illustration, that social scientists and government officials in Germany have been meeting to discuss the creation of public use samples. This meeting should be emulated by other countries.)

Some of the problems of use of social science methodological and policy research can be traced to the fact that researchers are not part of the policy formation activities of government. If social researchers are to play a greater role in social policy formation and are to increase utilization of their research, there must be a higher rate of communication between researchers and policymakers. This communication is more successful if social scientists participate in internal organizational decisions (14). Social scientists must make a concerted effort to involve themselves in these decisions. Involvement in the internal decision-making process will indirectly improve the quality of civil servants' activities and directly improve utilization of their research and policy recommendations.

With administrators and policymakers, social researchers can assess research needs and examine the relation of the statistical system to research activities outside the government. They can apply their training in organizational theory and public administration to improving information management activities in government. Indeed, some of these very activities are already underway in Italy, Norway, Germany, the United States, and Great Britain.

Closer ties between data producers and analysts will result in data that are more relevant to policy issues and will also improve the quality of both data and analyses. Producers of data will have more direct feedback on quality from major users of data...Users will come to have a better understanding of the operational problems of collecting and processing data, and will design and perform their analyses with a better understanding of the limitations of the data (11,168).

What should be the role of social science in an Information Age? This is a much more difficult question than the one which asks what knowledge should be applied and how? Let me identify only a few salient public policy issues that form part of an agenda for information technologies-related social research and training.

(1) Society will require a decreasing amount of work. Will work as a value lose its importance? How will the remaining work be distributed? What educational and job training programs will be needed, ones that are more compatible with the requirements of the post-industrial and information age? If the number of hours of leisure time is
increased, what social and psychological changes will occur; what changes will be necessary?

(2) New organizational structures are evolving and, increasingly, innovation takes place and new products develop in small units. What should be the role of the state in reorganizing the production structures? How do we design tax policies and write administrative regulations to provide incentives for industrial and university research and development, to foster innovation and risk-taking in the highly productive information technologies? If basic research outside industry is a prerequisite for innovation and continuing productivity, are the existing models of research in a more decentralized fashion, along the lines of the U.S. model, or research in the Colbertist tradition any longer relevant; or is some mix more appropriate to optimize available resources and to encourage innovation?

(3) Critical shortages of trained scientific and technical personnel are beginning to be felt. In what ways can we improve the quality of our science and social indicators to reflect the current situation? How can we estimate the impact of these shortages on the economy and on a nation's productivity? What roles should the state and the private sector play in ameliorating these conditions? If university budgets continue to experience serious erosion, how will a nation's productivity and general welfare be affected? Yet, if attention is turned only to reducing these shortages, do we risk neglecting the education of the "well-informed citizen" who is necessary for democratic control of technical decisions? Do we thus accelerate the creation of a society which is, to quote Shils, "victim of the parochial preoccupations of specialized technical experts"? (in McRae). If we emphasize scientific knowledge to the detriment of valuable discourse will we neglect the education of both the scientist and the consumer of technology?

(4) The design industry and regulatory arms of the state have been preoccupied with hardware systems, with minimal consideration of human factors and a disregard for worker participation. The accident at Three Mile Island nuclear power facility on March 28, 1979, dramatically illustrates the failure to integrate the reactor operator into the system. The Kemeny Commission pointed to the mutual isolation of the operator and equipment in the highly complex sociotechnical system as a root cause of the accident (15, 57). The social scientist Malcolm Brooks observed that the events were a direct function of the electro-mechanical system design and detail (15, 58).

In what ways can we improve the man-machine interface in order to reduce isolation and alienation? If it is necessary to modify the work environment, in what ways? Are our theories of participatory democracy relevant to the emergence of new environments based on information technologies? (Is the model of industrial democracy relevant in a post-Industrial Information Society?) Can the new information technologies and new sources of knowledge enhance autonomy and responsibility, make possible mastery of the natural and social world, and emancipate rather than imprison us?
(5) Instrumental reason has spread to many areas of social life and there is an increasing tendency to define practical problems as technical issues. Will technocratic domination erode the institutional framework of society? What value system will it dictate? Will the technical values of efficiency and economy dominate the selection of means for realizing social goals?

(6) The ability to communicate has always been the purview of the educated and dominant classes. Will standardization of access vocabularies affect language and syntax and authority structures? If language will be of a different nature, simplified, to reduce communication costs, will we then sacrifice part of the content? What will occur when the essential meaning of messages related to daily life becomes available to anybody? Will new communication structures create more open and accountable authority structures? Do they offer the potential of transforming the state into one more easily supervised by the "public"?

(7) The cultural model of a society also depends on its memory, control of which largely conditions the hierarchy of power. Will access to infinitely greater sources of information entail basic social changes and affect the social structures by modifying the procedures for acquiring knowledge? (10, 313). How will data banks restructure knowledge? How much social control will be exercised by the producers of data banks?

To understand the nature and direction of technological change demands a vigorous and sustained program of social research related to information technologies. The frameworks of the social science disciplines and social thought can help us in orienting our discourse and directing it to problems of action and choice. New information bases and new knowledge can improve political choices in an increasingly technological society. They can assist social groups to transform society, to use new resources effectively and to their benefit, and to create control mechanisms for the New Information Order. This effort requires engaging and appropriating competing traditions of philosophy and social thought, new philosophical approaches and different methodologies, and creativity and innovation unfettered by the narrow confines of the empirical sciences.

References


RECOMMENDATIONS FOR AN INTEGRATED DATA MANAGEMENT SYSTEM
FOR HISTORICAL SOCIAL RESEARCH
(SUMMARY)*

Whereas for social science research there now exists a wide variety of programs for statistical analysis, the software support for the specific demands of data management in historical social research is still insufficient.

Heterogeneity of sources, sequential data collection within archives, and reluctance to transform textual data into numerical codes very early, make it necessary to provide for more flexible instruments in the process of data collection and data management, that would prepare the data for subsequent statistical analysis through such widely distributed analysis packages as for example SPSS, BMDP, SAS, or OSIRIS. Those systems are nevertheless rather restricted as far as the processing of textual data is concerned.

As one consequence of this situation, researchers all over the world have started to create single purpose programs to bridge the gap between unstructured sources of research and the input requirements of statistical analysis packages. There are, notwithstanding, some approaches that try to integrate various of such problems into an overall system, such as SIR, CLIO, or even TUSTEP, a program for editing textual data.

Nevertheless, these developments have been created independently thus leading to isolated solutions, rarely compatible and machine-independent.

* These recommendations were devised by a working party established by the Center for Historical Social Research with financial support by Fritz-Thyssen-Foundation, Cologne. The full text has been published in HISTORICAL SOCIAL RESEARCH, No. 19, July 1981, pp. 83-92.
In historical social research it is often necessary to postpone the decision on how to transform textual data into numerical codes. This depends on the structure and duration of data collection within various archives. That is why historical social research needs more than what is currently provided in systems like SIR. Historical social research does not always need such special systems of data preparation. Especially if the sources are numerical in nature or when the transformation into numerical codes can be established by a small pretest, such software support would be irrelevant.

The working party favored the development of an integrated set of routines that should be obliged to the following main features:

- they should be machine-independent
- they should be embracing almost all known application problems
- they should be easily handled
- they should transform the non-numerical data for the direct input into statistical analysis packages
- there should be good instructions for the users

The set of routines to be used via a common meta-language should include:

- a high flexibility concerning transformation of different data types and structures (Input-Interface)
- a good many data-management functions (as e.g. data correction, data transformation, and record linkage, and postponed coding of textual data including the use of thesauri)
- some basic text processing (either to later allow for numerical assignments or to prepare textual data for text editing purposes)
- the preparation of data for subsequent statistical or graphical data processing.

The single functions that would have to be included in such a package would be:

- Ability to read a very vast number of logical data formats and structures. Beyond the usual fixed and free field formats, tag-content logic, hierarchical and network representations are a must.
- Data checking. The DDL components have to provide an easy
means to check the data being input for logical violations as, e.g., range violations, illegal strings, absence of structurally necessary variables and so on.

- Data modification and transformation. Beyond the usual simple possibilities for transformation as summing variables up, the user has to have access to automatically administrated codebooks and algorithms for complex coding.

- Record linkage. The major algorithms for the comparison of differently spelled names (not necessarily personal ones), which have been developed during the last years, have to be provided together with possibilities of logically adding the informational content of two data-bases.

- Data management and retrieval. The common requirements for DBMS's apply. The priority for flexibility of the command language and easiness of its use have to be incomparably higher, though, as the one for data security recovery of system errors.

- Basic features of textual analysis - tagging, use of stopwords - are required.

**NFAIS OFFERS INDEXING KIT**

The National Federation of Abstracting and Indexing Services, NFAIS, has announced an Indexing in Perspectives education kit, developed with partial funding from UNESCO. The cost is $30.

The purpose of the kit is to provide teaching aids that instruct librarians and information specialists in the development and use of indexes. Indexing in Perspective looks at the science of indexing from a historical and philosophical viewpoint, including insight into indexing techniques, the history of certain indexing systems, how indexes are arranged, the criteria for selecting an indexing format, and how to use indices most efficiently.

Designed for experienced teachers with a knowledge of indexing, the kit contains sections concerning indexing vocabularies, formats and retrieval, a glossary of indexing terms, lists of suggested workshops, the UNISIST Indexing Guidelines, and a bibliography. Supplementary teaching transparencies also may be ordered for $18. Contact:

NFAIS
112 S. 16th Street
Philadelphia, PA 19102

31
CLASSIFICATION ACTION GROUP REPORT

Mandate: At the last meeting of the CAG, it was decided not to change the wording of the mandate. It remains as is.

Past Year Report: The final version of discussions at the last Annual Conference was reported to the vice-president of IASSIST as requested and was subsequently published in the IASSIST Newsletter. There are no changes in that report.

Recent Activities: As a follow-up on one of the CAG Tasks, instructions on how to cite MRDF in the literature first appeared in a major social science journal -- SOCIAL FORCES. In the March 1981 issue, the "author's guide" section carried guidelines for citing MRDF. Subsequently, the same information will be incorporated into Carolyn Mullins' "A Guide to Writing and Publishing in the Social and Behavioral Sciences" to be issued as a new edition in March 1982. I have received reports that other journals in Canada and the US will follow the precedent set by SOCIAL FORCES.

Completing the work on the cataloging manual was another CAG Task. A draft of the manual was extensively reviewed in December 1980/January 1982. Recommendations from reviewers were incorporated into a revised version of the manual. The manual was then sent to three publishers (all of whom have expressed interest in publishing it). One of the three publishers requested some other revisions and these have been completed. It is expected that the manual will be published in 1982.

At the last annual meeting of the American Library Association Meeting in June of 1981, I represented IASSIST and the CAG at the Cataloging Committee: Description and Access session. At this session, I presented two reports: one was a series of recommendations to the committee based on concerns of catalogers of MRDF; the second was a report on how MRDF were represented in the International Standard Bibliographic Description (ISBD) for non-book materials. As a result of the first report, a Special Task Force was established to study the recommendations.
made in the report. The task force is made up of Alan Wajenberg (chair), University of Illinois; Elizabeth Herman, University of California at Los Angeles; Arlene Dowell, University of Chicago; Ann Fox, Library of Congress; and myself. The second report will be accepted as part of a five-year review of all the ISBDs. The Special Task Force on MRDF will meet in October 1981 in Washington, D.C. and will be reported back to the ALA Committee at the Mid-winter meetings in January 1982.

Another CAG-related task has been completed with the final version of the MARC format for MRDF. The CAG was to recommend data elements which could be used as access or retrieval points in an automated information system for MRDF. Both Sue Gavrel and I as respective co-chairs of the CAGs in Canada and U.S. were members of the working committee to establish a MARC format for MRDF. Carolyn Geda and Barbara Aldrich also serve on this committee. The MRDF/MARC format is near completion; a final meeting was held in October 1981.

The establishment and implementation of a 'cataloging-in-source' program has been achieved at ICPSR. This was another CAG Task. Cataloging information for both the MRDF and its documentation is provided on the verso of the title page at the time the documentation is issued to potential users.

A cataloging worksheet for MRDF has been completed and will appear in: Cataloging Machine-Readable Data Files: An Interpretive Manual. This completes five CAG Tasks. There is no change in the status of the three remaining tasks.

Sue A. Dodd, Chair
Classification Action Group, IASSIST

PERIODICALS ON MICROFICHE

Congressional Information Service, Inc., publisher of the American Statistics Index, has released the 1982 catalog CIS Periodicals on Microfiche, Backfiles and Current Year. This catalog lists and describes 253 important United States federal periodicals and publications that are available on microfiche from CIS. The backfiles date from the mid-1970's through 1981; current year subscriptions are available and fiche copies are sent automatically as issued.

For information contact:

Periodicals on Microfiche
Congressional Information Service, Inc.
P.O. Box 30056
Bethesda, MD 20814
(301) 654-1550 or toll-free (800) 638-8380
FID EDUCATIONAL SERVICES FOR INFORMATION WORKERS

The International Federation for Documentation (FID), Education and Training Committee has announced two projects. The Clearinghouse on Information Education and Training Materials, established in 1980, serves as a central source for materials and information useful for training in librarianship, information science, documentation, and archives work. Instructional support aids are solicited, collected and organized and distributed; aids include syllabi, reading lists, bibliographies, test problems, lecture notes, and teaching packages. Materials in the collection are from subject fields such as computer science, information science and documentation, reference work, information services, libraries and library science, systems analysis and others.

Materials can be forwarded to:

Clearinghouse
College of Library and Information Services
Hornbake Library
University of Maryland
College Park, MD 20742 USA

The Newsletter of Education and Training Programmes for Specialized Information Personnel, which began as an experiment in 1977, has been issued quarterly since 1979. It offers information on programs, activities, and educational developments in information science, documentation, library science, and archives. Highest priority is given to forthcoming events and recent contributions to the advancement of teaching and learning opportunities in the field. As the Newsletter depends on information received from educational institutions and personnel, and national planning and information agencies for its content, contributions such as press releases, program announcements, and new course details are solicited.

Please address all information to:

FID Newsletter on Education
c/o College of Library and Information Services
Hornbake Library, Room 1101
College Park, MD 20742 USA

Requests for copies should be addressed to:

FID
P.O. Box 30115, 2500GC
The Hague
Netherlands
Three catalogs of Canadian machine readable data files have been made available.

The Machine Readable Archives (MAR), a division of the Archives Branch of the Public Archives of Canada, is responsible for the collection, preservation, and servicing of machine readable records of historic value produced by the federal government and those of national significance created by the private sector.

Catalog of Holdings of MAR is the first in a planned series of publications to inform researchers of the machine readable files available; the catalog describes the files available through January 30, 1979. Prepared by Katharine Gavrel, with the assistance of the archivists, the catalog is divided into four sections: 1) descriptive entries, 2) title index, 3) principal investigator/organization index, and 4) subject index. Periodic updates of the catalog will be published.

Alcohol, Drug and Tobacco Use Files contain research data files which have been acquired upon request and in cooperation with Health and Welfare Canada. The booklet is an update of the 1978 publication Drug Use Files and contains a description of all data files within MAR relating to the use of addictive substances.

Both publications are available in English and French.

The University of British Columbia Library Data Library has introduced a computer output microform edition of that data library's catalog of machine readable files. This fiche edition supersedes all previous hardcopy editions and contains: 1) an introduction to the University of British Columbia Data Library's collections and services, 2) descriptions of all data files in the collection (as of November 1981), 3) an outline of the numeric subject classification codes used in the catalog, and 4) alphabetic title and principal author indices.

A product of the public SPIRES database, this catalog can be searched interactively with a University of British Columbia identification number, locally at the university or remotely through DATAPAC.

For further information, contact:

Data Library
University of British Columbia
6356 Agricultural Road, Room 206
University Campus
Vancouver, British Columbia, Canada
V6T 1W5
THREE NEW REFERENCE BOOKS

Gale Research Company has announced publication of three new reference sources.

The first edition of the International Research Centers Directory, edited by Anthony T. Kruzas and Kay Gill, was published in November 1981. Consisting of three paper-bound volumes, the directory covers 1500 research organizations throughout the world; arranged by country, entries include university-related, government, and independent research organizations. It is designed as a companion to Gale's Research Centers Directory and the Inventory of Major Research Facilities in Europe, published by K. G. Saur and distributed by Gale in the Western Hemisphere.

Edited by P. William Filby with Mary K. Meyer, Passenger and Immigration Lists Index is a guide to published arrival lists of nearly 500,000 passengers who emigrated to the United States and Canada in the 17th, 18th, and 19th centuries. Typical main entries include: 1) name and age of passenger, 2) date and port of arrival, 3) code indicating the specific source which contains the arrival record and the page number within that source, and 4) names, ages, and relationships of any accompanying passengers. Cross references for accompanying passengers to the main entry are provided.

The 1982, seventh edition of Statistics Sources, edited by Paul Wasserman and Jacqueline O'Brien, is a subject guide to data on industry, business, social, educational, financial and other topics, both national and international. This volume contains citations for nearly every country in the world, with increased coverage of the Soviet bloc. Arranged in a dictionary style with frequent cross references, Statistics Sources cites publications compiled by trade and professional societies, local, state and federal government agencies, foreign governments and international bodies. The principle statistical sources for each country are identified; an additional feature is the "Selected Bibliography of Statistic Sources" which is an annotated list of important English language general statistical compendiums.
INDEX, VOLUMES 3-5

Administrative Committee, IASSIST 3:1:18
3:2:41-43
4:1:3-7
Aldrich, Barbara
"Approaches to Information About Census
Bureau Machine-Readable Data Files" 4:1:3-7
Association of Public Data Users (APDU) 4:2:50
Avedon, Elliot 3:2:45-46

Beck, Carl
Behavioral and Social Sciences Librarian 4:1:20-21
Brosveet, Jarle 5:4:73-78
Brown, Christopher D. 3:2:30-37
Brunken, Erika von 3:2:23-29
"By the Seat of My Pants" - see Robbin, Alice

Census Microdata, 1976 3:4:34
Census '80: Continuing the Factfinder Tradition 4:2:49
Census, 1980, Seminars 3:3:70
Civilizations Course Materials Project (CCMP) 5:4:85-86
Classification Action Group (CAG), IASSIST 3:2:44
3:2:47-48 4:2:47-48

Clearinghouse on Education and Training Materials 4:2:50-51
Committee of European Social Science 3:4:31-33
"A Computer Center Survey: What Students Don't Like" 5:3:56-63
"The Conduct of User Surveys" 4:3:64:66-68
Congressional Information Service 4:1:12-15

Danish Data Archives 3:1:19-20
Data Acquisition Group, IASSIST 3:2:44-45 4:2:45
Data Archive Development Action Group, IASSIST 3:2:46-47
4:2:45
Data Archive Registry Action Group, IASSIST 3:2:45-46 4:2:45
"Data Archives Retrospective and Perspective" 3:4:6-13
Data Center for Housing Research 4:3:64:76-77
Data Organization and Management, IASSIST 3:2:43-44
4:2:45

Data Protection and Social Science Research:
Perspectives from Ten Countries 4:1:19-20
Data Services: Development and Management Interest
Group (DSDM), IASSIST 4:2:45
5:3:64-65
"Data Services for the Uninitiated" 5:1:25
DeVries, John 3:2:46, 48-49
Dhanak, Harriet A. 3:2:30-37
Directory of Data Holdings and Services (Social
Science Data Archive, Yale University) 4:2:48
Drug Abuse Epidemiology Data Center (DAEDAC) 5:1:23

37
Education Committee, IASSIST
European Cooperation in Social Sciences Information and Documentation (ECSSID)

Feasibility of Maintaining and Providing Access to Data Archives Through ERIC
Fisher, Gary P.
Gammell, William
Gavrel, Sue
General Social Survey, 1980

Harrison, Donald
Heim, Kathleen M.
Henrichsen, Bjørn
Holm, Lars W.
Housing Research Data Center - see Data Center for Housing Research

IASSIST Committee and Group Reports

IASSIST Committee Structure

IASSIST Conference, 1979
IASSIST Conference, 1980
IASSIST-IFDO Conference, 1981
IASSIST, Proposed Constitutional Changes
IASSIST, Treasurer's Report, 1979
IASSIST, Treasurer's Report, 1980
IASSIST, West Coast Group

International Demographic Data Center
International Federation of Data Organizations (IFDO)

Kasputys, Joseph E.
Knoppers, Jake
Kolp, John G.
Lacasse, Pierre
"A Library-Based Reference Service for Machine-Readable Census Data: The Canadian Experience"
3:2:44-45
3:3:55-65

McDonald, Dennis D.
Machine-Readable Data Files: A MARC Format
4:3&4:66-68
3:4:33-34
4:1:15

Manojlovich, Slavko
Mochmann, Ekkehard
"A Model for User's Service: Providing for Information and Data Retrieval from an Archival, User and Development Library"
3:2:30-37
3:3:55-65
4:1:19

Monkkonen, Eric H.
Muller, Paul
5:2:28
5:1:23-24
3:1:19-20
3:4:28-31

Nijhawan, N.K.
"Notes on the Social Sciences as Producers of Technologies"
5:3:49-55

"On-line or On Tape"
"On-line Reference Tools for the Hard Sciences"
"The Organization and Management of the Norwegian Social Science Data Services (NSD)"
5:4:79-84
4:2:38-40
5:4:73-78

"Peace Research and Information Systems"
"President's Report on 1981 IASSIST-IFDO Conference"
"Private and Public Sector Responsibility for the Collection, Distribution and Analysis of Statistical Data"
Process-Produced Data Action Group (IASSIST)
3:1:3-13
4:3&4:57-59
4:3&4:69-72
3:2:46
3:3:66-67
5:4:3-22

"Recent Historical Studies of Crime and Crime Control in the United States"
"Reference Tools for Machine-Readable Data Files"
"The Revised IASSIST Constitution"
Robbin, Alice
3:4:14-22
4:2:33-38
5:2:28-41
3:1:14-16
3:2:38-41
3:3:66
3:4:23-28
4:1:23-24
4:2:52-53
4:3&4:57-59
5:1:3-22
5:4:87
5:3:49-55
3:3:66
4:2:41
3:2:48
5:4:79-84

Roberts, Barbara
Rockwell, Richard C.
Rokkan, Stein
Rowe, Judith
Ruus, Laine G.M. 3:2:46-47
Sande, Terje 5:4:73-78
Sidell, Philip S. 4:3&4:78
Sobel, Richard 5:2:44-47
"Social Trend Studies: A Review Essay" 5:2:44-47
Software Review 5:4:86
"State and Regional Data Archives (SRDA)" 4:1:7-10
Statistical Reference Index 4:1:12-15
Stein Rokkan Prize 4:2:41
"A Strategy for Archiving Government Data to Meet the Needs of the Research Community" 4:3&4:60-66
Strauss, Murray A. 4:1:7-10

Tessier, Judith A. 4:1:22
Trees, Don 4:3&4:74-76

Urban Indicators: A Guide to Information Sources 5:4:87
"User Needs and Confidentiality in Sweden" 3:2:23-29
"User Services in a Data Library" 4:2:29-33
USSR-USA Working Group on Social Science Information 3:1:18-19

West Coast Local Seminar Survey 4:3&4:74-76
Wood, Gordon H. 4:2:38-40

Zentrum fur Historische Soczialforschung e.v. quantumforschungsstelle 3:3:67-68

Indexed by Carolyn Latta