

FINAL REPORT OF THE
JOINT WORKING GROUP OF THE
SOCIAL SCIENCES AND HUMANITIES RESEARCH COUNCIL
AND
STATISTICS CANADA
ON THE ADVANCEMENT OF RESEARCH
USING SOCIAL STATISTICS

DECEMBER 1998

EXECUTIVE SUMMARY

Canada's economy and society are in a period of rapid and difficult change. Timely and objective analysis of economic and social conditions is required to understand this transformation, to provide a basis for broad and informed debate on public policy, and to establish a foundation for intelligent policy formation. The need is particularly acute because Canada's social policy has not kept pace with the dramatic changes in its economic policy over the past two decades. Governments at all levels have acknowledged the importance of redesigning social policy so that it meets the needs of all Canadians, and leads us towards more civil and economically sustainable communities.

In one respect, Canada is well-equipped to meet these needs. We now have a number of excellent and timely social surveys covering a variety of topics. Theoretical developments and advances in research design have led to the creation of longitudinal surveys which track individuals over extended periods of time. These new research tools provide information on the dynamics of poverty, the effectiveness of training programs, the consequences of job loss, the influence of childhood experiences, and several other topics pertinent to redesigning social policy. Together, these form the basis for establishing a well-integrated system of "social statistics", a term we use to encompass information describing a wide range of human activity and the social, economic, educational, and cultural features that affect our daily lives.

However, as a nation we have very little capacity to conduct social policy research, evaluate social programs, or monitor progress towards achieving social aims. The Federal Government recently acknowledged the need to strengthen its research capacity and established the Policy Research Initiative to recommend and oversee the implementation of an interdepartmental research agenda. Similarly, provincial governments, the private sector, and non-government organisations recognized this need and have attempted to revitalize the policy research community.

There are at least three significant barriers that need to be overcome if we are to develop our research capacity in social statistics. The first is the sheer lack of trained researchers. More than a decade of Federal and Provincial Government restructuring and downsizing has significantly reduced the number of researchers working within government institutions. This has occurred during a period of decline in the training of statistics and research methods in social science departments of our universities, with the exception of economics. The second significant barrier is a lack of access to data. Paradoxically, the detail provided by new data sets, which makes them so valuable, prevents them from being made public, because of their potential to identify specific individuals. The Statistics Act sets out strict criteria for maintaining confidentiality, and Statistics Canada necessarily has a very strong position on ensuring the law is followed. The third barrier is the very weak links between the work of social scientists and the potential users of the knowledge they generate. Even though there is a tremendous appetite for social statistics about education, employment, health, literacy, and other pertinent social issues, many of the important findings of social scientific research have not been adequately conveyed to the policy community, or to a wider public through the popular media.

The mandate of our working group, convened jointly by Dr. Ivan Fellegi, the Chief Statistician, and Dr. Marc Renaud, President of the Social Sciences and Humanities Research Council (SSHRC), is to make proposals to encourage quantitative research on major social and economic issues using large-scale data. This report recommends the funding of three components, each designed primarily to surmount the barriers described above. Together, they comprise a **Social Statistics Research System**, which would complement the integrated system of social statistical data for which Canada is already considered a world leader.

The first component is aimed at increasing the number of researchers engaged in quantitative research on social and economic issues. It includes three separate proposals: **Research and Training Groups** that would bring together researchers from different disciplines and institutions to conduct quantitative research and training in priority areas; a **Training Program**, including a summer school, that would provide specialized training in advanced statistical methods complementary to apprenticeships and graduate programs and support for data librarians to increase the use of social statistics in undergraduate training; and a **Fellowship Program**, including M.A., PhD and Postdoctoral Fellowships aimed at providing support for young researchers pursuing careers in social statistics, as well as Senior Fellowships that would enable some of Canada's leading social scientists to devote more of their time to research, and provide leadership in the training of the next generation of researchers.

The second component entails two proposals: **Research Data Centres** and remote access capabilities

that would provide access to detailed micro-data, while maintaining the strict rules for the preservation of confidentiality required of Statistics Canada under the Statistics Act; and support for the enhancement and expansion of the activities of the **Data Liberation Initiative**.

The third component is the development of a **Social Statistics Communication Program** that would implement a communications strategy to inform and build the public constituency for quantitative social science research. A key component of this program are **Research Forums** that would support research networks, provide an arena for the presentation of research findings, and enhance communication among researchers, the policy community, and the media.

The report also recommends that Statistics Canada and SSHRC negotiate a memorandum of understanding that defines the goals and organization of the proposed Social Statistics Research System. A co-ordination structure is also recommended, so as to take advantage of opportunities to strengthen networks and increase synergy among researchers.

Table of Contents

	PAGE
I. Introduction	1
II. Building Research Capacity	3
A. Research and Training Groups	3
B. Improving Skills: Training Program and Summer School.....	5
C. Fellowship Programs	6
MA, PhD and Postdoctoral Fellowships.....	6
Senior Fellowships in Social Statistics.....	6
III. Access to Data	7
A. Research Data Centres	7
Remote Access	9
Seeking Approval from Respondents to Share the Data with Data Centre Users	9
B. Archiving and Documenting Data	10
C. The Data Liberation Initiative	10
IV. Communicating Research Findings	11
A. Research Forums	12
B. Social Statistics Communications Program.....	13
V. Co-ordination, Costs and Timetable.....	14
A. Co-ordinating the Canadian Social Statistics Research System	14
B. Costs and Timetable.....	15
Appendix A – Some Priority Areas for Research in Social Statistics.....	16
Appendix B – Curriculum for a Summer School Program	21
Appendix C – Research Data Centres	23
Appendix D – The Development of a Remote Access Capability at Statistics Canada.....	28
Appendix E – A Model for Research Forums in Social Statistics: the Canadian Employment Research Forum.....	29
Appendix F – Membership of the Joint Working Group.....	30

I. INTRODUCTION

Over the last decades, Canada has experienced - and continues to experience - dramatic social, economic, and technological changes. Our economic policy has responded to and adjusted to these changes for a number of reasons. In particular, however, the adjustment has been in response both to pressures stemming from global international markets, and to the need to abate inflation and reduce the national debt. These changes in economic policy have been accompanied by major changes in the labour market and in family structure. There is a general sense among many Canadians that the major problems we face are not economic, but social. Governments at all levels have acknowledged the need to redesign our social policy so that it fits better with our current economic policy.

To meet this need we require research on a wide range of social, economic, educational, and cultural issues. The renovation of social policy must build on a basic understanding of the life course, and the complexity of social relationships at different levels, such as families, neighbourhoods, and communities. Conducting such research requires a well-integrated system of social statistical surveys, together with the capacity to analyse the data.

Beginning with data from the census and surveys on trade, finances, prices, the labour force and other areas, Statistics Canada has developed a national statistical system. Historically its aim has been to provide aggregate indicators of current economic and social conditions, including descriptive statistics on economic growth, health, education, justice, productivity, the labour force, and so on. During the 1970s, two major changes shaped its evolution. First, research in the social sciences demonstrated that an understanding of many social phenomena - such as the nature of criminal activity and victimization, or the effects of poverty - required separate, focused surveys. Second, the policy community recognized the importance of research that could help us understand how particular life events, combined with peoples' habits and lifestyles affected their long-term social outcomes. Research began to stress the importance of *social context* - the families, neighbourhoods, schools, and organizations in which people live and work - in shaping and constraining individuals' actions. But the descriptive data available from cross-sectional surveys were inadequate for monitoring changes in social outcomes, or understanding the causal mechanisms that led to desirable outcomes. This required longitudinal surveys, in which data were collected from the same sample of respondents on at least two, and preferably more occasions.

Statistics Canada responded to this challenge by instituting a new generation of social surveys during the 1980s and early 1990s. They include, among others, the General Social Survey (GSS), the Graduate Follow-up Surveys, the Survey of Labour and Income Dynamics (SLID), the National Longitudinal Survey of Children and Youth (NLSCY), the National Population Health Survey (NPHS), and the Displaced Worker Survey (DWS). Over the next few years, additional major new surveys are planned, including the longitudinal Workplace and Employee Survey (WES), the longitudinal Youth in Transition Survey (YITS) and the Survey of Financial Security (SFS). Meanwhile, Statistics Canada also strengthened its efforts in collecting administrative data and taxation data, and its capabilities to link data from various sources. Canadian researchers also played a major role in the development and administration of international surveys, such as the International Adult Literacy Study (IALS) and the Third International Study of Mathematics and to Science (TISMS). Several other surveys are currently being developed, and efforts are being made to integrate the survey data so that they can better characterize Canadian life. Together, these surveys have taken the national statistical system in a new direction, and have helped distinguish Canada as a world leader in the collection of social statistical data.

Despite this great strength, Canada has relatively little capacity to analyse the data from these surveys. The most difficult barrier to surmount is that there are simply too few researchers engaged in quantitative research. The problem is especially acute in areas requiring advanced statistical methods. The capacity to conduct research within government institutions has declined over the past two decades, mainly due to restructuring and downsizing. Although these institutions appreciate the need to bring analyses of social statistics to bear on public policy debates, they are not strongly enough connected to the universities, where the majority of social scientists conduct their research, and where the training of new researchers occurs. Within the universities, training in statistics and research methods has also declined severely during this period. There are now very few faculty who teach courses in advanced statistical methods, and in most social science fields, the course requirements for M.A. or PhD degrees no longer include formal training in statistics or quantitative research methods. The notable exception is economics, but in this field research

capacity has been and continues to be eroded by the migration of qualified researchers to the United States.

This problem is exacerbated by the fact that the data generated by the new surveys are relatively complex: the surveys are usually longitudinal, and have a multilevel structure, such as students nested within schools, or workers within firms. Thus their analysis requires powerful computing and statistical techniques. There have been tremendous theoretical developments in this area, but there are very few social scientists in Canada who have training in their application.

Another significant barrier is access to data. In the past, much useful research was conducted on public use data files made available to university and other researchers by Statistics Canada, especially after the launching of the Data Liberation Initiative (DLI) by the academic community, with support from Statistics Canada, the Humanities and Social Sciences Federation of Canada, and SSHRC. A difficult problem is posed by the need to make detailed micro-data – the exact responses to questions in surveys – available to researchers, without compromising the confidentiality promised to survey respondents. The powerful statistical techniques which are appropriate for the analysis of multilevel, longitudinal data cannot be performed with aggregate data; access to micro-data is essential. Moreover, even simple descriptive problems often require access to micro-data. Part of the problem, ironically, is that the detail on individuals provided by longitudinal micro-data could make it possible for a researcher to identify specific individuals. This would violate requirements to maintain the confidentiality of individuals' responses, as set out in the Statistics Act.

The third barrier to the development of our research capacity in the social sciences concerns communication. Many social scientists necessarily devote their energy to addressing narrowly defined research questions that will advance their academic discipline. Success in academic careers depends to a large extent on publication in scholarly journals, and usually there are relatively few incentives or resources available for preparing more popular pieces that would convey research findings to a broad audience. Moreover, the demands of an academic career to excel in research and teaching make it difficult for researchers to forge strong links with the policy community, and often the time span of academic research is too long to meet the immediate needs of policy-makers. Consequently, the transfer of knowledge from research to policy and practice is not as rapid or as strong as it could be. Perhaps for the same reasons, social scientists have not been particularly successful in popularizing their findings through the public media.

However, in the longer term, the strengthening of social science research requires widespread public recognition of the benefits of social policy research. We need to demonstrate that empirical research provides a basis for informing the development of public policy, evaluating social programs, and monitoring our progress towards achieving social aims. There is no substitute for a specialized dialogue among researchers and policy-makers. As this will not occur spontaneously, an organized effort is required to increase public awareness of research findings.

A Joint Working Group of the Social Sciences and Humanities Research Council (SSHRC) and Statistics Canada was convened by Dr. Ivan Fellegi, the Chief Statistician, and Dr. Marc Renaud, President of SSHRC. Its mandate is to present a set of proposals to encourage quantitative research on major social and economic issues using large-scale data. The effort of the Group takes place in the context of other initiatives aimed at improving research using social statistics. The Federal Government recently established the Policy Research Initiative (PRI), with a mandate to recommend and oversee the implementation of an interdepartmental research agenda designed to remedy gaps in the knowledge required for policy development in the Federal Government. The PRI has partnered with SSHRC to undertake a project on societal trends, with the objective of gaining insight into major social changes, to understand their implications for policy research, and to identify key areas for future research. In addition, SSHRC is developing programs of strategic research in the areas of social cohesion, the knowledge-based economy and society, and population health. These are also priority research areas for the PRI.

This report recommends the funding of three components, designed primarily to increase the number of researchers engaged in quantitative social science research, improve access to micro-data without compromising confidentiality, and inform and build the public constituency for social statistics. Together, these components comprise a **Social Statistics Research System**, which would complement the national statistical system which is already well developed. These components are described separately in the next three sections of this report. The final section sets out a time-line and estimated budget for implementing the proposals, and recommends a process for co-ordinating the inter-related activities that would follow.

II. Building RESEARCH CAPACITY

The first component of our proposed Social Statistics Research System aims to increase the number of researchers engaged in quantitative social science research, and to improve the skills of those already working in this field. It contains three separate proposals: **Research and Training Groups**, a **Training Program**, including a summer school, and a **Fellowships Program** that includes M.A., PhD, Postdoctoral, and Senior Research Fellowships. These are described below.

A. RESEARCH AND TRAINING GROUPS

We recommend establishing a program of Research and Training Groups with four interrelated functions: the conduct of quantitative social research; the training of researchers; the broad dissemination of research findings; and the provision of feedback about the usability of data sources. We envision medium-size research groups, preferably cross-university and cross-disciplinary, working on major research problems, focused on data from a number of different sources. The program for Research and Training Groups would have an adjudicating committee, including researchers from abroad, that would implement a strategic, multi-stage, and pro-active adjudication process.

Central to the development of a strong and diversified body of research is a community of scholars who work together and encourage and criticize each others' work. Research communities that focus on particular research issues are especially critical for graduate students and researchers beginning their careers, as they enable them to accumulate experience and judgement which is not easily learned from books or through university courses. SSHRC has experience with a variety of different kinds of organizational forms, and it has found that the most productive environments are those where researchers meet face-to-face and work on a common intellectual theme.

The centrepiece of our initiative, therefore, is a program to support research and training groups. We envisage middle-sized groups, typically with 8 to 15 researchers. This number provides a critical mass, while avoiding the coordination problems of larger groups. Each Research and Training Group would focus on a topic that is sufficiently broad to merit a co-ordinated program of research, training, and dissemination activities; and sufficiently important to attract a significant number of researchers. Ideally the Groups would be multi-disciplinary, and entail partnerships among disciplines within a university, and between two or more universities. These Groups would focus their efforts on research pertaining to a particular theme.

Appendix A (see p. 16) of this report describes some promising themes where there is already an abundance of available data.

There are already social research groups in Canada, in areas such as aging, education, immigration, and child development, with funding through SSHRC's Major Collaborative Research Initiatives (MCRI) and National Centres of Excellence. The proposed Research and Training Groups, however, would focus on quantitative social science research, and place much more emphasis on developing the skills of new researchers and enhancing the skills of an entire research team. Good research teams provide valuable informal learning as a by-product of conversations, meetings, and formal seminars, but the Research and Training Groups would be required to address the need for training more explicitly. They would identify

training needs and describe concrete steps to address them. They might create and improve formal courses, or prepare background research papers on research methods or aspects of survey design. Efforts to establish apprenticeships that would bridge the gap between formal graduate courses and applied research would be of primary importance.

The Groups would also be expected to develop and carry out plans for the dissemination of their research results, not only to specialized audiences of academics, but to policy makers and the general public. This would involve more than an episodic flurry of communications activity; it would entail sustained efforts at reaching out, culminating, in several cases, in the formation of partnerships with groups outside academia. Finally, these Research and Training Groups would provide feedback to those collecting data, in particular to Statistics Canada, about the quality of data being collected, its relevance to central research problems, and its limitations. We expect that this would improve future data collection efforts, and sustain researchers' commitments to this activity.

Research and Training Groups that work in similar or related areas would, over time, become involved in virtual networks – regionally, nationally, and internationally. These networks would allow researchers who were not geographically close to interact with the Groups, or even to become members. Research and Training Groups would be encouraged to pursue their training, dissemination, and feedback objectives by putting to use other components of the Canadian Social Statistics Research System that we propose: the Summer School and the specialized workshops to be offered by the Training Centre, the Research Forums, which would offer the opportunity of meeting with a variety of research users, and the Communication activities to be developed at SSHRC.

We expect that most of the Research and Training Groups would focus on the analysis of large-scale Statistics Canada surveys. SSHRC and Statistics Canada would encourage research that uses complex longitudinal and multi-level data, and comparative research, especially using multi-country databases. Research and Training Groups might also plan and conduct their own surveys, either free-standing, such as studies of Canadian elections, or surveys that involved Canadian participation in an international social survey. Conducting such surveys outside of Statistics Canada, or in partnership with it, would increase opportunities for training students how to plan, design, and carry out a social survey. This would benefit Statistics Canada and the survey research industry.

Applications for Research and Training Groups would be peer reviewed, with successful applications funded for an initial period of 5 years, and renewable. Typically, the project budgets would support student research assistants and postdoctoral fellows, computers and software, travel, and in some cases, data collection and archiving. Funding could also be granted to support non-student research professionals; in particular these Groups would likely require professionals capable of preparing and documenting the complex datafiles required to conduct sophisticated quantitative analysis, of providing ongoing support to graduate students who undertake data analysis, and of solving problems with computers and software.

We recommend that the Research and Training Groups program have a single, separate adjudicating committee, with international members who have expert knowledge of quantitative social science research, including policy-related research. The committee would be charged with developing an adjudication process that was comprehensive and formative; that is, one that encouraged the development of new proposals and links with other groups and the policy community. We feel a separate adjudication committee is necessary as it would be able to understand the special requirements of quantitative social sciences research, and balance the criteria, preferences and objectives to be served by the Research and Training Groups. It would also serve to situate the activity of the Groups within the larger context of the Canadian Social Statistics Research System, and attend to the on-going development of Groups, especially when they are in their early stages of their development.

B. IMPROVING Skills: TRAINING PROGRAM AND SUMMER SCHOOL

We recommend the establishment of a Training Program, including a national Summer School, covering topics in quantitative methods and the use of Canadian micro-data. The program would serve policy analysts, university researchers, and graduate and undergraduate students. It would also provide support to data librarians.

To complement the Research and Training Groups, there is a need for a program of short courses in social data analysis, similar to the American and British efforts at Ann Arbor and Essex, but more decentralized and adapted to Canadian data and research concerns. The goal is not to offer a substitute for university courses at the graduate and undergraduate levels, but to provide a series of intermediate-level courses in data analysis, and more advanced courses pertaining to the analysis of longitudinal and multilevel data. These courses would stress the application of statistical techniques, and would use Canadian micro-data.

A Training Program of this nature is required for several reasons. One is that students at the senior undergraduate and graduate levels who take training in statistics are often unable to make the leap from the theoretical training offered in university courses to applying the techniques to real data. In many cases, the examples offered in textbooks pertain to social issues in the U.S., and employ small, contrived data sets. Many students need some basic tutoring on how to manage data, such as how to retrieve data from a CD-ROM, merge files, select cases, and sort data. They also tend to need instruction on rather simple topics which are seldom covered in their training, such as scaling variables, creating composite variables, handling missing data, and using design weights. Some basic courses that used one or two Canadian data sets could generate considerable enthusiasm for pursuing a career in social statistics. Also, when students take such courses alongside other university researchers and policy analysts, they learn first-hand about the kinds of problems that the analysis of micro-data can address. A considerable amount of incidental learning occurs, and important working relationships are formed, many of them inter-disciplinary.

Researchers who are at an advanced level also require training in the more complex statistical methods appropriate for analysing longitudinal and multilevel data sets. This is essential if we are to exploit the richness of the recently developed surveys. The audience for these courses would be mainly university researchers, and policy analysts working in government agencies and other organizations. These courses also need to be provided centrally, because there is not usually a critical mass of researchers wishing to pursue such advanced training at any one university.

Finally, because data librarians play an essential role in facilitating access to and use of the existing databases, efforts should be pursued to increase their training. This would contribute to undergraduate education as well, because data librarians can help professors use Canadian data in their courses, and assist students in getting started on projects that require some data analysis.

The logical centrepiece of an effort to improve training in quantitative data analysis in Canada is a summer program, modelled somewhat after the ongoing and successful summer programs held at Michigan and Essex. The idea is to offer a series of courses in one or more locations, normally over a three- to four-week period. Ideas for a curriculum are provided in **Appendix B** (see p. 21). However, this program would differ from the UK and U.S. programs in three important respects: first, it would focus on Canadian social issues and use Canadian data; second, it would provide a social and intellectual environment for graduate students, faculty members and public sector researchers with an interest in quantitative research on Canadian issues; third, it would provide a means for follow-up activities, such as establishing an electronic network and the exchange of research papers. It would also give researchers an opportunity to establish contacts with personnel at Statistics Canada who are directly involved in managing particular data bases.

There is no reason to restrict these activities to the Summer. The Training Centre should provide a platform for related educational and consultative activities throughout the year. These could include workshops to develop plans for new surveys, seminars to support research on a given topic or with a given data-set, and workshops on statistical techniques and survey methods. Departments, foundations, businesses, unions, and NGOs could use the facility for consultations and seminars on pressing research questions. Activities of this kind already take place on an *ad hoc* basis, but a Centre with a mandate to

facilitate such training would dramatically reduce the effort presently required for researchers to conduct such training programs.

We propose to begin with a program organized by Statistics Canada and using their teaching facilities in Ottawa. Aside from providing a good, ready-made teaching environment, Statistics Canada's ability to operate both in French and in English is extremely attractive. From the beginning, Statistics Canada would work closely with academics through an advisory committee. As the program becomes established, some of the courses could be taught at various universities. In the longer term, the program itself could be moved to a university base in one or more universities, while retaining a partnership with Statistics Canada. Also, there should be an effort to reach an accreditation agreement with some universities, as this would likely be helpful for many students.

We recommend that funding be provided to initiate the program. Funding is required for the salary of a director, office expenses, travel, instructors, and travel and honoraria of the advisory committee. The main source of funding for the Centre would be tuition fees and sponsorships for specific activities, such as workshops on a particular data-set. In the longer term, this program should strive towards becoming self-financing. Attracting graduate students to a summer program will require SSHRC support for travel, accommodation and tuition. Support of this kind might also be made available to junior faculty members. It would be appropriate to approach MRC and NSERC, who could support graduate students in areas that would benefit from this program.

C. FELLOWSHIP PROGRAMS

We recommend that SSHRC establish a targeted program to support students and young researchers at the M.A., PhD and Postdoctoral levels in social statistics, and a program of Senior Research Fellowships in Social Statistics.

To develop a stronger research community in the longer term, we must ensure that young scholars are attracted to pursue a career in social statistics. The field needs a distinct and higher profile, visible to graduate students and post-docs as they decide on the orientation of their research. For this purpose, they should have access to specifically targeted financial support.

M.A., PhD and Postdoctoral Fellowships. The fellowships program would be distinct from the regular SSHRC fellowships programs, but its basic eligibility conditions, regulations and the size of awards would be the same. Applications would be restricted to quantitative social research, though the requirements for demonstrating this will vary according to the level of the award. M.A. candidates would be expected to take graduate courses that provide them with theoretical and analytic skills, and to write a review paper or thesis on a quantitative topic in social statistics. The application for PhD and Postdoctoral Fellowships would include a proposal to conduct a quantitative research project.

Given the targeted nature of this program, evaluation of these applications requires the establishment of a separate expert adjudication committee. Also, there should be an explicit commitment to achieving a gender balance. We therefore recommend that all applications for these fellowships be adjudicated directly at SSHRC.

Senior Fellowships in Social Statistics. The committee recommends that *two* Senior Fellowships be granted every other year to university scholars who are making a significant contribution to research and training in the area of Canadian social statistics. The Fellowships would be for \$100,000 per year for a period of five years, and would require that recipients devote at least 80% of their time to research and training. Renewal would depend on the recipient maintaining a high level of scholarly research. The Fellowships would be prestigious, and would likely increase the momentum of the scholars' academic careers. These fellowships are modelled after those awarded by the Canadian Institute for Advanced Research (CIAR). The CIAR fellowships have been highly successful in keeping some of our best scholars in Canada, and enabling them to develop significant research programs.

We recommend the adjudication committee be the same as for the other fellowships proposed above.

Scholars could be nominated for a Senior Fellowship by their university, or they could apply directly to SSHRC. The application would entail a brief description of how their proposed research program would contribute to social statistics and to the training of young scholars. Adjudication would be based on the applicants' records of scholarly activity, and their proposed research and training programs.

Internships. We have also identified a need for more young researchers to obtain experience in policy research, in view, among other things, of eventually finding employment for their newly developed skills in quantitative social sciences. Although we see the immense potential benefit of a program of internships, we could not develop precise ideas in this respect in the absence of a more general program of internships at SSHRC. We can therefore only recommend that the new internship program proposed in the Council's Innovation Scenario take into account the special need for social statistics expertise identified in this report, and that appropriate resources be provided for it. Moreover, we think that in view of the development of the Social Statistics Research System being proposed here, Statistics Canada should more fully promote its own internship programs, and encourage other government departments to do the same.

III. ACCESS TO DATA

If Canadians are to fully benefit from the substantial data resources that exist, researchers and analysts must have adequate access to these data. Research is required to convert data into usable information, and adequate access is necessary to facilitate and promote research. Data access has many dimensions. It can mean ensuring that users have adequate knowledge of and access to basic statistics from the data sources. It can also mean access to public use micro-data files required for more complex analysis, and at yet another level, access to the detailed unscreened (confidential) micro-data. Archiving is also an access issue, as data sources may not be accessible in the future if care is not taken to properly archive them when they are created.

The Data Liberation Initiative (DLI) significantly improved the access to public-use household survey data sources. The Joint Working Group viewed this initiative as extremely positive, and recommendations are made in this section regarding ways in which the initiative might be extended. Recommendations are also made regarding support for the archiving of data created by research groups and initiatives outside of Statistics Canada.

The major issue dealt with in this section, however, relates to the access by researchers to data which are confidential under the Statistics Act. The principal issue is how to create such access while ensuring that the respondents' confidentiality is protected as required under the Statistics Act. Confidentiality is the cornerstone of the statistical system. It is also true that in order to fully exploit the new data sources being created, researchers need to have some form of access to the detailed data. A number of alternative means of achieving these goals are discussed here, but the main proposal relates to the creation of Research Data Centres

A. RESEARCH DATA CENTRES

The committee recommends that SSHRC and Statistics Canada jointly create a national system of Research Data Centres, where researchers can access detailed micro-data for research purposes, while maintaining the confidentiality provisions of the Statistics Act.

Increasingly researchers require access to detailed micro-data to conduct research in many areas, including health, child development, income and economic security, labour adjustment and workplace change. Paradoxically, as the quality and complexity of the data have increased, access to these micro-data may decline, as a result of confidentiality requirements. Confidentiality protection is crucial. Not only is there a commitment of confidentiality of responses made to citizens when they provide information, but the security of data is vital to maintaining the public and political confidence required to carry on the broad range of data collection activities effectively.

Since the early 1970s, Statistics Canada has made micro-data (individual records) available while protecting confidentiality through the production of "public use" micro-data files for household surveys. To prevent the release of records in which the information from individual respondents can be identified, some data must be suppressed: this may involve the omission of some elements of the individuals' records or "collapsing" response categories, so it is not possible to identify respondents with unusual characteristics. But this solution is no longer adequate for two reasons. First, many of the new longitudinal surveys have a data structure that prevents public-use micro-data files from being created at all, for it becomes increasingly difficult to develop a micro-data file that maintains confidentiality when longitudinal data are available for a number of years. Second, researchers with more complex theories and more powerful statistical tools increasingly find that the suppression of information in traditional public use files significantly limits their analysis. If the data access issue is not addressed, not only will a number of new and innovative data sources be underutilized and valuable research lost, but researchers may increasingly turn to data from other countries, particularly from the U.S., where the access problem has been more adequately addressed. Findings from American research data are often not applicable to the Canadian situation.

One means of addressing this issue is to have researchers sworn in under the Statistics Act as "deemed employees" of Statistics Canada, allowing them to conduct research using the detailed micro-data files. The Statistics Act provides for such deemed employee status under conditions that are discussed later. Even after being sworn in under the Act, however, there is still the issue of physical access to the data. It is proposed that a number of locations be created across Canada where such access would be possible. These offices would be called **Research Data Centres**, and would be modelled on a similar program established by the U.S. Census Bureau. Following is an overview of the Research Data Centre proposal. The details of the proposal can be found in **Appendix C**. (see p. 23)

Ultimately, we envision several research data centres distributed across major urban centres and university campuses throughout the country. A centre would provide a secure physical location for the confidential data. Legally it would be an office of Statistics Canada, and a Statistics Canada employee would be on site. The centres would operate a research program administered by SSHRC and Statistics Canada. Researchers who had a proposal accepted under the program, and were sworn in under the Statistics Act, would have access to confidential micro-data files maintained at the Centres. A committee that operated under the auspices of SSHRC and Statistics Canada would carry out the adjudication of the program submissions, which would be on the basis of merit. The Selection and Review Committee(s) would consist largely of academic researchers, with some representation from other research communities and Statistics Canada. The research merits of the proposal would be paramount in the selection process (combined with resource issues). Researchers at universities, research institutes, government agencies, and other research organizations could submit proposals.

The Statistics Act governs access to confidential data, and its conditions must be incorporated into any program of Research Data Centres. Researchers may become deemed employees of Statistics Canada and access confidential data providing that they are "perform(ing) special services for the Minister (i.e. Statistics Canada)." Since Statistics Canada clearly does not have the human resources necessary to conduct all the research necessary to exploit the data sources, arranging for other researchers to provide such service is one means of achieving this goal. In this environment the research done would be similar to that which Statistics Canada itself would normally conduct. While this may sound restrictive, in fact Statistics Canada carries out a wide range of quantitative research; therefore, this is not likely to be an issue.

The output of the research program associated with the data centres would consist, in the first instance, of research papers produced by the researchers. The papers would undergo the normal peer review process, which would be managed by the Selection and Review Committee. It is anticipated that the vast majority would then be included in the Research Paper Series associated with the research program. Since the research program would be run jointly by SSHRC and Statistics Canada, these papers could not comment directly on policies or programs, as that is outside the mandate of Statistics Canada (Appendix C on p. 23 provides more detail.) Beyond this first stage, the researchers would be free to publish a revised paper (or the original research paper) in an academic journal, or any other publication. There would be no concern in this case as to whether the product contained policy comment; the authors would be free to include any material they deemed necessary. It is proposed that the copyright for the original research paper would be vested with the researchers. Statistics Canada would retain the right to publish the original research paper, and to vet all publications for confidentiality.

As long as the site had a level of security comparable to that maintained by Statistics Canada, a Research Data Centre could be located at a university or a non-governmental research institution. For convenience, however, some Centres could be attached to existing Statistics Canada regional offices. The advantage of non-Statistics Canada locations is that they would allow the institution serving as a home base to develop a very strong empirical research capacity. Empirically-oriented researchers would be attracted to an institution with a Data Centre, allowing the university or institute to develop a strong program. However, it may be appropriate to begin this program with a pilot project at a Statistics Canada regional office. It is proposed that Research Data Centres initially include surveys and administrative datafiles with the exception of the ones concerning businesses; issues arise in the access to the latter that would need to be addressed before they were included in the program.

There are at least two alternative funding approaches for research data centres. The first, modelled on the Data Liberation Initiative (DLI), is for the Centres to be entirely “block funded”, with the costs shared by some combination of SSHRC, Statistics Canada, the university or institute at which the centre is located, and other organizations, such as government departments and research institutes. If, as we anticipate, primarily academic researchers would use the centres, the majority of the block funding would be SSHRC-based. In this model, there would be no direct cost to the researchers using the centres. The selection process, conducted by the selection and review committee, rather than applications for funding, would regulate access to the centres. A variant of this approach would be to have part of the annual cost of the Centres covered by block funding, and the remainder covered by fees paid by researchers using the centre. University-based researchers whose projects were approved by the committee would receive a SSHRC grant, distributed through the review committee to cover this cost; researchers from other organizations would pay similar fees. In any case, care should be taken to facilitate access to data for graduate students and post-docs.

Remote Access. In addition to deemed employees accessing confidential data at Research Data Centres, there are at least two complementary potential solutions to the data access problem. One is to develop a remote access capability in Statistics Canada. Researchers would submit their computer jobs to the agency, and they would be run on confidential micro-data files by Statistics Canada staff. The output would be vetted by Statistics Canada staff to ensure that confidential data were not released. This approach is complementary to, and not a substitute for, the Research Data centres. It would be useful for researchers with smaller projects, or those who were not physically close to the Research Data Centres. The remote access would not, of course, allow the flexibility and quick interaction required by many research projects that the Research Data Centres would provide. A Statistics Canada committee has been struck to develop a proposal in this area, and an outline of their proposal is in **Appendix D**. (see p. 28)

Seeking Approval from Respondents to Share the Data with Data Centre Users. The Statistics Act includes a provision for the sharing of data with users. If the respondent's permission is sought at the time of data collection, confidential micro-data may be shared with selected users. Data sharing would be possible with an Institute created by SSHRC, or some other body. Statistics Canada would then provide the raw micro-data (for respondents who agreed to share, most of whom do) to the Institute, which would of course agree under a contract arrangement to maintain confidentiality.

This approach is quite straightforward. The selection and vetting process would then be entirely in the hands of the Institute; Statistics Canada would play little role other than setting up the conditions to ensure data confidentiality. The shortcoming of this approach is that it can only be applied to data collected in the future, and to some data sets. The data sets that have been created to date could not fall under such an arrangement. For that reason, this Joint Working Group believes that the approach should be seriously considered for the future, but will not solve the current issues. In the longer run this approach is a potentially excellent solution to sharing selective data sets, and should be seriously considered once the initiatives in this report have reached some level of maturity.

B. ARCHIVING AND DOCUMENTING DATA

We recommend that SSHRC allow researchers who created data files before documentation and archiving was an allowable expense in SSHRC grants to apply for funding to conduct such documentation and archiving on the more important older data files. We also recommend that SSHRC and/or DLI consider the creation of a national archiving system for quantitative social science data sets created outside of Statistics Canada.

Organizations other than Statistics Canada often produce large quantitative research databases. SSHRC-funded initiatives can generate such files. The more important of these data sources should be archived for future use, and the data bases should be created and maintained so that they are readily accessible and transportable. If this is not done, researchers other than those initially involved with the data have no access to it, particularly some time after the file was created. There is currently a lack of support for archiving important research databases created some time ago. In SSHRC-funded projects involving the collection of data, the cost of providing documentation and creating files for dissemination purposes has become an allowable expense. The difficulty is with databases assembled without SSHRC support or with only partial support over a period of time and those that pre-date the funding allowances for archiving. Examples of such data files include The Aging in Manitoba data set, The Canadian Study of Health and Aging, and The Canadian Fertility Studies. Our understanding is that at this time, it is not possible to apply for funding to place the data in a form that can be disseminated. We therefore recommend that SSHRC expand one of its existing programs or develop a new program to provide for such funding. Applicants would have to make the case that the data are not presently in a form that allows them to be easily distributed or archived, provide plans for preparation of the data, a description of the final products and demonstrate that the data are of sufficient importance and interest that they will actually be used if archived and disseminated.

C. THE DATA LIBERATION INITIATIVE

We recommend that the Data Liberation Initiative be invited to submit proposals with a view to pursuing and extending its mission of making data and research information easily available for academic research and for undergraduate training; this could include the training of data librarians, and the extension of networks of co-operation among them and with undergraduate instructors, as well as the preparation, dissemination and exchange of teaching materials.

The Data Liberation Initiative (DLI) was launched a few years ago by the Humanities and Social Sciences Federation and Statistics Canada, together with a consortium of post-secondary institutions and federal departments. It has been an immense success, allowing researchers in universities and colleges much easier access to public use data files. Many more researchers are developing an interest in quantitative data. A new generation of data librarians have been appointed and trained in most institutions, and they have learned to work in close co-operation with one another to increase the availability of both data and the information required to analyze them. Prior to the DLI, data centres existed in very few institutions, and now they are being created in many more. Software has been developed to provide meta-data on available data sources, and to help with their extraction: students and relatively inexperienced researchers can now easily and rapidly bring quantitative data to bear on their analytical ideas; this nurtures their initial interest, and paves the way for further involvement with social statistics.

Throughout this period, many data librarians have also engaged in collaborations with social science instructors who teach undergraduate students the logic, methods, and practice of data analysis. Many of these instructors have incorporated in their teaching much more hands-on experience with data by students. Moreover, they can send their students, graduate as well as undergraduate, to the data librarians to get help in accessing the data and the accompanying documentation.

The Data Liberation Initiative now is under review as it approaches the end of its initial five-year phase. It is a free-standing organization, with its own Board and administration, and it is not the role of our Joint Working Group to make their case in their stead. This being said, we need to say how important it is that they be supported by Statistics Canada and SSHRC, morally as well as financially. The original mission of disseminating public use data to academic researchers should of course be preserved; but other roles that have been assumed by the DLI should be recognized, pursued and augmented. First, the training of data librarians, who have become indispensable intermediaries between data sources and users, should be expanded, and the standards of the profession should be sanctioned by employers and employees alike. Also, the networks of co-operation among them should be further extended and strengthened.

Second, the DLI is ideally placed to assume a larger and more systematic role in improving undergraduate education in methods and social statistics, in co-operation with post-secondary institutions, as well as with the Humanities and Social Sciences Federation, Statistics Canada, and SSHRC. It can promote exchanges among undergraduate instructors, and with data librarians, about improved ways of teaching, through written and electronic exchanges as well as in workshops, for instance at the annual Congress of the Social Sciences. It can also improve the availability of teaching materials for these instructors, in the form of raw or semi-processed data, documentation about data and concepts, articles and research reports, and suggestions for course assignments.

Third, the work of the DLI, centrally as well as in the local data centres and in networks, may be seen as the precursor for an eventual Canadian data archive. Such archives already exist in most advanced countries, and they have proved invaluable research instruments. While the creation of such an archive at the national level is not presently planned, that may change, and the work of the DLI in producing better meta-data and documentation is most likely to move us in that direction. Therefore, it should be strongly supported. Indeed, the proximity of data librarians to the users will help to keep the material readily usable by the latter. We can anticipate that university-based data centres will even succeed in preserving and disseminating the precious procedural information developed in the course of using quantitative data in research. In other words, a data archive might grow from the bottom up over the first few years of maturation of the Canadian Social Statistics Research System; SSHRC and Statistics Canada should pay attention to potential developments, and be prepared to lend them support.

IV. COMMUNICATING RESEARCH FINDINGS

Ultimately, support for an active program of quantitative social analysis depends on public support and interest. Indeed, supplying relevant research and bringing it to the attention of people who could use it in debate and decision-making is key to increasing demand for such research. But at best, the process of recruitment and incentives around academic research does little to reward the abilities and temperament required to arouse public interest. With notable exceptions, policy-makers find contemporary social science either not directly relevant to their problems, or too narrowly focused. And researchers, in turn, are often reluctant to engage in work that appears unduly oriented towards short-term practicalities. The links between social science research and the media are also relatively weak, in spite of the tremendous appetite of the public for social statistics about education, employment, health, social assistance, aging and other social issues. In both cases, obstacles crystallise, in particular, around the dispersion of the information, the inaccessibility of the language, and differences in operational timetables.

While some progress has been achieved in certain research areas, we need to be more proactive in the promotion of dialogue between social scientists and the potential users of their findings. What is needed is a change in culture, which can only come about after a rather extended period of interaction; and such interaction will in turn require that opportunities be systematically created and exploited for this purpose. This is why we propose the creation of Research Forums, as well as the organization of a Social Statistics Communications Program at SSHRC.

A. RESEARCH FORUMS

We propose the establishment of a program for the support of research forums in order to bring together social researchers based in universities, governments and non-governmental research organizations with a broad range of policy analysts and decision-makers. The forums would sponsor conferences and related activities at which academic and government researchers would debate, and would benefit from each other's research, data, and experience.

Too often today, academic research on social and economic questions of interest to policymakers in federal, provincial and municipal governments is never seen by those potential audiences. At the same time, academic researchers remain unaware of emerging policy debates to which their expertise can usefully contribute. As a result, a huge pool of expertise for the greater understanding of Canadian social trends and policy issues remains untapped. To bridge these gaps, we propose a program of Research Forums.

Our model for Research Forums is the Canadian Employment Research Forum (CERF), which was formed in 1991 to foster Canadian research and bring together policy researchers in Ottawa and academic labour economists. As recently as ten years ago, empirically-oriented Canadian labour economists worked primarily with U.S. data. Many Canadian faculty members and graduate students knew more about the workings and evaluation of the U.S. welfare and unemployment insurance systems than their Canadian equivalents. At a very modest cost, CERF has effectively and rapidly eliminated this problem, and has created a vibrant, interactive community of academic and government researchers. More information about CERF's activities is provided in **Appendix E**. (see p. 29)

We believe that CERF's formula can be adopted in other areas of inquiry, and applied to a broader range of activities designed to bring together researchers and policy analysts from a wide range of institutions. A successful forum requires a critical mass of researchers in a well-defined but not too specialized area of research. The main activity of the forums would be to organize conferences, with the goals of informing policy makers and policy analysts of research results, providing researchers with a better understanding of policy debates, and more generally offering a setting to facilitate contact among the different sectors of the Canadian Social Statistics Research System. We also see this as an important opportunity for bringing graduate students and beginning researchers into a research milieu. Forums could also provide a convenient and effective vehicle whereby researchers in both the academic and government sector could discuss advantages and limitations of Statistics Canada's surveys for addressing current policy questions, and transmit these insights to Statistics Canada.

We propose that SSHRC conduct peer-reviewed competitions for research forums in quantitative social science. Each forum would be funded for a period of five years, and would be renewable as long as the objectives of the program were met. We envisage the creation of one or two forums per year in the initial five years. For simplicity, these applications could be assessed by the same committee that assesses Research and Training Groups. To qualify as a forum, a group of researchers would form a steering committee, of approximately a dozen members, who would be co-applicants. In addition to the university-based researchers, the appointment of researchers from government and the non-profit sector to the steering committee should be encouraged. Participation and, where possible, financial support from other organizations, including universities, government departments and the non-governmental organizations concerned with public issues, should be viewed positively in assessing applications.

Applications for new forums should include a plan for an inaugural conference in the first year, plus an indication of its direction for the following years. Of course, existing groups with activities of the kind we have described may apply for support. The funding is mainly required for a forum's key activity – organizing conferences – and will include some secretarial and logistic support, travel (both for university faculty and graduate students) and small honoraria for the preparation of conference papers that involve a major departure from a researcher's ongoing research program. To circulate conference papers and maintain communications, funds should cover the maintenance of a web site; and some support should also be

available for publicizing the conferences. Because a successful forum will depend heavily on the largely volunteer efforts of academics in organizing conferences, building and running the organization, some teaching release for Canadian academic members of the steering committee should be available. Travel funding is also required for attendance at meetings of the steering committee.

B. SOCIAL STATISTICS COMMUNICATIONS PROGRAM

We recommend that SSHRC establish a Social Statistics Communications Program, specifically aimed at increasing public awareness of quantitative social science research. In co-operation with Research and Training Groups, with Research Forums, and with Senior Fellows, and in consultation with the communication services of universities and of Statistics Canada, this Program should work towards raising the profile of social statistics research findings, and towards furthering debate around the publication of such statistics.

Social statistics are a privileged area for communications in the social sciences, because of the steady stream of quantitative information about social and economic phenomena being published in the media. Various sections of the public, as well as non-governmental organizations, have also come to expect these numbers and the debates that accompany their announcement. Given this favorable context, social statistics should become a priority in the communications strategy of SSHRC. Their Social Statistics Communications Program should be articulated around two streams of information, one coming out of research in academia, the other from the regular publications of Statistics Canada.

In the first place, university researchers produce a number of research findings that are of great interest to the public and to policy-makers, and this will increase as Research and Training Groups come into being. The paradox here is that communications are best managed centrally, or at least from a few key locations where the information can easily be made available to all who could use it, while the research supported by SSHRC's programs is produced in a multitude of research groups and centres, distributed in a number of institutions of higher learning. The challenge, then, is to gather and organize the information, and to prepare researchers for its effective transmission, while seldom having direct control over the way in which communication does take place.

The Communications Program should first monitor the flow of research involving social statistics, using for this purpose connections to Research and Training Groups, to Senior Fellows in Social Statistics and, above all, to Research Forums. Indeed, the Program should be systematically represented on the Boards of the Research Forums. Most of these can be expected to have web pages, and there should be a central web page of these web pages. The media's attention should also be drawn to interesting and relevant research being produced or presented in these places; a bulletin could be published regularly, including for each item the name of reference persons that can be contacted about the findings. Moreover, the Program should have regular relationships with the communications services of universities, encouraging them to raise the profile of social statistics research. This could be achieved both by having universities signal such research to the Program, and by the latter providing support, when needed, to these university services. Finally, the Program could offer training workshops to researchers, for instance at the Congress of Social Sciences, on how to deal with the media.

In the second place, Statistics Canada regularly publishes a variety of social and economic statistics, thus bringing issues to public attention. When the profile of specific issues has thus been raised, SSHRC could take advantage of the moment to make the ideas of academic social scientists more visible to the public and the media. While the purposes and operations of the Communications Program at SSHRC will remain quite distinct from those of the Communications Division at Statistics Canada, some co-operation would be helpful. For instance, given some advance notice about Statistics Canada's publication calendar, the Communications Program could present the media with a current list of experts in any given area of research, and help identify potential contributors to debates on social statistics. Given the decentralization of many aspects of social policy, attention could be paid to identifying experts coming from the various

regions of Canada. The Program could itself organize presentations, and debates among experts, on issues where social statistics can be used to throw light on issues. The public for these events would often be the media, but it could also be public servants, politicians or various organizations. Co-operation and a division of labor with the Humanities and Social Sciences Federation would be useful in this regard.

Finally, the Program could, in co-operation with the Research Forums, organize lecture tours and workshops where research findings based on quantitative social science analysis would be presented to various groups, especially to potential new researchers in social statistics in universities. This activity could be undertaken in co-operation with Statistics Canada, which is already active in this field.

V. CO-ORDINATION, COSTS AND TIMETABLE

A. CO-ORDINATING THE CANADIAN SOCIAL STATISTICS RESEARCH SYSTEM

In order to co-ordinate the implementation of the program in this report, and to monitor progress, we recommend that SSHRC and Statistics Canada agree to a memorandum of understanding specifying their short- and long-term objectives under this initiative. The two organizations should also jointly appoint a co-ordinator for the overall Research System and an advisory committee of senior officials from academic and other organizations.

The initiative in quantitative social research we have described involves an ensemble of interrelated programs. A strong research environment involves a set of mutually-reinforcing institutions, which is why we refer to an emergent “Canadian Social Statistics Research *System*”. The ongoing coordination of various components could benefit from an explicit agreement on goals and the formal coordination of the program elements. The actors filling the different roles in the system will need more systematic information about each other than informal networks can provide. There are also some overarching issues that do not fall within the mandate of any one of the components of the system, such as efforts to foster intellectual and methodological perspectives that cut across all research areas and activities. We recommend that this goal-setting and coordination take the form, first, of a memorandum of understanding between SSHRC and Statistics Canada, and, second, of a position for a co-ordinator of the initiative.

We recommend the appointment of a full time co-ordinator, at the director level. The co-ordinator should have extensive experience in research brokerage and management, as well as the intellectual breadth and research experience required to communicate effectively with a range of social scientists. Part of the co-ordinator’s mandate should be to seek out and develop partnerships that will facilitate the overall research effort.

The co-ordinator could convene a liaison group, involving representatives from all the components of the system : the chairs and the administrative officers of the peer review committees dealing with the Research and Training Group, fellowships and forums programs; representatives of the Research and Training Groups and Research Forums themselves; co-ordinators of the Summer Workshops and Research Data Centres; a representative of the communications program. The co-ordinator would report to the Chief Statistician and the President of the Social Sciences and Humanities Research Council.

In order to build and maintain strong linkages to the academic and policy research communities, it is also recommended that a senior level advisory committee be formed. Representatives on such a committee should include a number of senior representatives from the academic community (e.g. Vice-presidents for research) and government (e.g. Assistant Deputy Ministers for policy), as well as international researchers or research managers from countries that have successfully implemented similar research initiatives.

B. COSTS AND TIMETABLE

The proposed research initiative consists of a number of inter-related components. The Joint Working Group did an initial costing of the various components by considering the costs of similar programs. We recognized that a much more extensive costing exercise would be required as the proposals are transformed into detailed activity work plans.

We also recognized that the full program will take a number of years to reach full development, and the assumption has been made that the program would be phased in over a five-year period beginning in fiscal year 1999-2000. The Working Group feels there is a pressing need to proceed quickly on a number of components, and that work could and should begin in this next fiscal year. The highest priority should be given to the start-up, at least on a pilot basis, of several Research Data Centres. It should also be possible to initiate the training program by beginning with a few selected courses, and to launch a few research forums by holding a number of research conferences on priority topic areas. Work should also begin at SSHRC on a communications strategy with respect to social statistics. It is recognized that major competitive programs, in particular the Research and Training Groups and Senior Fellowships, will require a longer review process; they should consequently be started as soon as possible.

Appendix A

SOME PRIORITY AREAS FOR RESEARCH IN SOCIAL STATISTICS

A wide range of policy issues and research areas could benefit from the strengthening of quantitative research which would put to use the available databases. The role of the Joint Working Group, however, is not to “pick winners,” but to establish programs with fair selection processes. The brief topic descriptions which follow are intended to suggest, but not restrict, the research that this initiative would stimulate.

1. CHILD DEVELOPMENT

In the past ten years, there has been an increased awareness that the quality of children’s experiences during the formative years has long-term effects on their happiness and well-being, their future educational and occupational experiences, and their health status as adults.

A research agenda on early childhood development in Canada could focus initially on the following questions:

- What is the prevalence of Canadian children who are vulnerable to unduly negative life experiences stemming from poverty, family violence, inadequate parenting, or racial and ethnic prejudice?
- To what extent is childhood vulnerability related to family structure, especially single versus two-parent families, and socioeconomic factors, such as family income and parents’ education?
- In what ways do the answers to these first two questions depend on the age of the child, and the cohort?
- What are the buffering mechanisms or protective factors associated with healthy child development?

As the policy community, including parents, teachers, administrators, and government policy-makers, attempts to design a new social policy for Canada, it must figure out ways to strengthen families and communities without dramatically increasing government expenditures. Among practitioners, there is a sense that clinical interventions for all children at-risk are too costly and do not adequately meet the needs of all of them. But are interventions targeted for particular groups a better alternative, or would universal programs likely have a stronger impact? If so, what types of programs would be most effective? Research on child development could provide a means of monitoring our progress towards reducing childhood vulnerability and redressing inequalities.

To address these questions, the chief source of data is the National Longitudinal Survey of Children and Youth (NLSCY).

2. YOUTH IN TRANSITION

The transition from high school to post-secondary education and from education into the labour market is problematic for many youth. While the research literature identifies the family background factors mainly associated with poor academic and occupational attainment, considerable research is necessary to understand the pathways to success and critical transitions for youth between the ages of 15 and 25. Some of the principal questions that might guide a program of research include:

- Which groups are particularly prone to leaving school before graduation, and have the most difficulty in making the transition to post-secondary education or to the labor market?
- What are the skills, attitudes and behaviors of youth who achieve successful transitions and those who do not?
- To what extent do graduation rates vary among schools within each province? What school-level factors contribute to successful graduation rates and to high academic achievement?
- Do programs such as co-operative education, mentoring programs, distance education and apprenticeships help students make these transitions?

At least five national and international data sets can be brought to bear on these questions: the National Longitudinal Survey of Children and Youth (NLSCY), the Program of International Student Assessment, the planned Canadian Youth in Transition Survey, the International Literacy data, and the Education and Training Surveys.

3. FAMILIES IN FLUX

Over the last three decades, family life has changed profoundly. The increasing proportion of women in the labour force (especially mothers with young children), as well as the growth of flexible and atypical employment have modified the gender division of labour, both within families and within society, and led to a reorganization of family time. The rise of separation and divorce, the decline of marriage, and the increase in cohabitation have transformed the family trajectories of women, men and children. Researchers have just begun to understand the far-reaching implications of these changes. The questions to be answered include:

- What are the effects of the changing labor environment on the propensity of men and women to both start and maintain conjugal and family relations?
- How are the existing relations among family members (e.g. between conjugal partners, between parents and children, between children and grand-parents) modified and redefined once the family separates?
- What are the consequences of family disruption on fatherhood?
- How are resources shared between partners once a union is dissolved?

To address most of these questions, longitudinal data are required. Such data sets include the 1984 Family Survey, the 1990 and 1995 General Social Surveys on the Family, the National Longitudinal Survey on Children and Youth, and the Survey of Labour and Income Dynamics.

4. GROWING OLD IN CANADA

In recent years, researchers, policy makers and program managers have begun to direct attention to the process of aging and to the status of being old in Canada. Major gaps in our knowledge about growing old in Canada include:

- Good descriptions of the uneven retirement process of Canadians from the age of 50. Canadian policies and research are largely based on the misconception that everyone in the labor force enters at about 20 and retires at age 65.
- Knowledge about the oldest old, aged 85 and over. Important questions concern the health of this group, care giving and receiving, living arrangements and income security of people with limited employment-based pensions (including CPP/RRQ).
- The characteristics of people who will likely enter and are currently in residential institutions. Knowledge about these groups is critical to health, housing and economic security planning and policies.

Numerous data sets exist to address these issues, including the National Population Health Survey, Survey of Labour and Income Dynamics, General Social Survey cycles on related topics, CARNET data, the Longitudinal Administrative Data files, the Canadian Study of Health and Aging, the Survey of Ageing and Independence, the Health and Activity Limitation Survey, and the Residential Care Survey.

5. EDUCATION, SKILLS AND LITERACY

The development of skills and human capital, along with technological advancement (both “soft” and “hard”) are seen as the primary forces driving productivity, and hence the standard of living in modern economies. It is difficult to overstate the importance of education, training and skill development for most societies. While this topic has been the focus of major research efforts, rapid changes in the economy and society, and particularly changes in the role and significance of education and skills requires on-going research. In a “knowledge-based” society, the following issues are currently on the policy agenda:

- The performance of Canadian students in a national and international context
- The effect of changes in the education system on access to higher education

- Lifelong learning and its implications
- The role of literacy, *independent* of educational attainment, in labor market success and daily activities
- The adequacy of training in Canadian firms
- The link between human capital and technological change
- Skills shortages and oversupply
- The role of human capital in wealth development.

A number of existing data sources cover these issues, including Adult Education and Training Surveys, International Literacy Surveys, Academic Achievement Tests, Graduate Follow-up Surveys, the planned Workplace and Employee Survey, and traditional surveys such as the Labor Force Survey as well as censuses.

6. THE DISTRIBUTION OF WAGES AND WORK

Developed economies today face central policy issues concerning changes in the distribution of wages and of work. In particular, many countries, including Canada, the U.S. and Britain, have seen a substantial increase in wage inequality in the last twenty years. Key research questions include:

- What has caused these trends: increased imports from low-wage countries, and the related phenomenon of outsourcing to those countries, or the introduction of new technologies that eliminate the jobs of less skilled workers?
- Is the collapse of unskilled men's wages a result of the declining influence of unions? Do changes in the quality of our basic education system play a role? Does the relative supply of highly educated workers influence inequality? If more than one of these factors is at work, what is their relative importance and the pattern of their interplay?
- Does a country's institutional structure affect whether it has an unemployment problem (like France and Germany) or a wage inequality problem (the U.S.), or both (Canada)? Does labor market policy respond to research findings, and if so, how?
- What will the role of social support systems be in the future?

Substantial research has been conducted on these topics, but many questions remain. New data sources, such as the Survey of Labour and Income Dynamics and the Workplace and Employee Survey, as well as more traditional sources such as the monthly Labour Force Survey will shed new light on the role of skill-based technological changes and related questions.

7. SOCIAL AND COMMUNITY SUPPORTS

Over the past decade there has been an increasing recognition of the importance of unpaid work activities, including childcare, household work, eldercare and volunteer work. Of course, women continue to do the majority of unpaid work, despite the increase in their paid labour force participation. The Federal Government's Policy Research Initiative has identified changing time allocation over the life course and within each stage of the life course as underlying many of today's social policy challenges. Care for both children and seniors is also undergoing changes. Aging of the Canadian population coupled with shifts in the responsibility for care, away from institutions and towards individuals and families, are major challenges. The combination of a declining age of retirement with increased life expectancy may also result in time imbalances at older ages. Questions are increasingly being raised about the erosion of community support or "social capital".

Data addressing these issues include: material in various years of the General Social Survey dealing with time use, and with social support and caregiving; the National Survey of Volunteering and Giving; the national Censuses; the National Population Health Survey; the Canadian Study of Health and Aging and the National Longitudinal Survey of Children and Youth.

8. SOCIAL IMPACTS OF SCIENCE AND TECHNOLOGY ON FAMILIES/CHILDREN AND ON WELL BEING

Science and technology are dominating forces in this century, perhaps even *the* dominating forces. They are often argued to benefit quality of life and children's futures, and yet we know little about the longer-term social impacts of technological change. It is difficult to make broad statements about whether their effects are positive or, more generally, to describe how they work. Also, there are poorly understood distributional issues. Who benefits and who loses; for example, what is the long-term effect of the gap between children with school- and home-based access to computers and the Internet?

Little is known about the social impacts of science and technology on families and children, and on well-being. Current efforts by Statistics Canada to construct a coherent framework for the systematic development of statistical information for science and technology present opportunities for analytical exploitation of existing data and the production of new data vehicles and linkages.

A number of data sets exist, or will soon be available, to address these issues, such as the Innovation Surveys, the Workplace and Employee Survey, the Graduates Surveys, special surveys on Internet use, research capabilities, and so on. The development of new data is also necessary to extend research areas.

9. EVOLVING WORKPLACE AND TECHNOLOGY USE

The 1990s has seen the intersection of a number of technology-related phenomena that affect the workplace and workers, among which:

- Increased use of information technologies, including rise of the Internet and related communications technologies in almost all industries, accompanied by a concern about whether this has led to the desired productivity gains
- Increased focus on the importance of innovation for firm survival and growth, and for productivity gains
- The effect of technology on downsizing
- Concern that technology may be leading to increasing polarization in society
- A focus on human resources issues such as training, pay practices, work schedules, and new workplace practices implemented to achieve "high performance" workplaces.

Relatively little is known about the adoption and diffusion of technology and innovation in work organizations and its effects on the workplace and workers. While researchers have long been concerned with these issues, few large-scale data sources have existed to document the rate of implementation of technology and innovation, let alone its implications. This has led to the use of often-questionable proxies for technology use, or case studies.

More recently new bodies of data have evolved to assist in our understanding of issues in this area, and to provide new research opportunities. Such surveys include technology surveys, innovation surveys, the Workplace and Employee survey, and a survey of the determinants of firm growth.

10. WELFARE, INCOME AND POVERTY

Research on welfare and material inequality addresses three main questions. First, what is the distribution among individuals, families, household units and "communities" of income, wealth and the necessities of shelter and food? Second, what are the personal consequences of this inequality on the quality of social life, generally and on specific issues such as health? The third question is what mechanisms reproduce, or alter, inequality over time, both over the lifetime of individuals and between generations? At the aggregate level, this question can be reframed in terms of the evolution of geographical and other forms of communities through time?

Studies of welfare, income and poverty are necessary to understand the effects of a very wide range of policies involving huge expenditures including: the redistributive effects of taxation, the efficacy of social welfare programs, education and health programs, and arrangements for the delivery of social services and health care.

Basic profiles of economic and social inequality have been available for several decades. But as policies and programs change, new research is required to assess their impacts. The new longitudinal surveys provide a hitherto unavailable way to examine closely the process of change over time. The impact on the welfare of Canadians of changes such as the loss of a job or the dissolution of a marital union can be addressed. The short and long-term impact of persisting conditions of severe deprivation on the well-being of children and young people, and on their physical and mental health can be assessed.

Data available to address these issues include the Survey of Consumer Finances, the Survey of Labour and Income Dynamics, the National Longitudinal Survey of Children and Youth, the National Population Health Survey, the Family Expenditures Surveys, and censuses.

Appendix B

CURRICULUM FOR A SUMMER SCHOOL PROGRAM

While it may take several years to reach a steady state, it is useful to think about the curriculum of a more mature summer program, distributed over four weeks. For data analysis, the core of the program should be two courses, each two weeks long. The first would provide a basic introduction to regression, designed for participants with relatively little background, with data analysis examples from cross-sectional surveys such as the General Social Survey. A second two-week course, in linear models, would extend regression techniques to censored, categorical, "count" and duration data. The two courses should combine lectures in the morning with mentored practice sessions in the afternoons. In this and other courses, there should be an emphasis, and specific instruction, on how to write about the results of data analysis.

In addition to the two core data analysis courses, on a revolving basis and according to the demand, each summer program should include one or more intermediate-level courses on topics such as longitudinal analysis, hierarchical models, latent variable models, and categorical data. The emphasis would be on building a corps of researchers equipped with modern data analysis techniques to conduct state-of-the-art research. It will often be appropriate to focus on a single survey, especially for longitudinal analysis, where considerable effort is required to gain familiarity with a data set. There should be a mixture of one- and two-week courses, depending on the topic. Some shorter workshops could address, on a rotating basis, more limited analysis topics such as weighting and estimation issues for complex samples, the analysis of pooled cross-sections, and robust statistical methods.

A third kind of methodological course, definitely worth trying, was proposed to the Working Group by two statisticians (one at Statistics Canada and the other university-based). The teaching groups would focus directly on a data set, chosen on the basis of teachers' and students' interests. Students would work as a team, led by a researcher and including a programmer, a survey methodologist, a specialist on the survey topic and a statistician. The course would revolve around using the chosen data to address a specific research problem, with design, statistical and substantive issues addressed as they arose.

Consideration should be given to staging a workshop, perhaps a week in length, on the philosophy, logic and strategy of data analysis. In recent years, there has been more interest in trying to make explicit, and consequently debatable, the thinking underlying the quantitative analysis of social data. For example, a considerable debate over the relative roles of description and causal interpretation of social phenomena was sparked by the publication of Stanley Lieberson's "Making It Count". Such a course might combine an introduction to the problem of what and how much particular data and data analysis can tell us about social processes, with a discussion of strategies for analysis, focussing particularly on complex survey data sets and on how to do analysis that does justice to a complex topic, but is still feasible.

In addition to the applied statistics courses, each summer session should include at least one course geared to developing analysis of a single survey, such as Survey of Labour and Income Dynamics, the General Social Survey on a topic, the National Population Health Survey, or the National Longitudinal Survey of Children and Youth. These could be open to researchers prepared to make a commitment to conduct a piece of research on the relevant data set and should be led by two or more researchers working with the data set; they would also involve an explicit mentorship arrangement for junior researchers. More senior researchers would take the course mainly to familiarize themselves with a data set and might collaborate in mentoring. According to the topic, the data-familiarization should be paired with a statistical workshop dealing with methods required to analyze the data. Where appropriate there might be follow-up activities, including a strategy for providing ongoing methodological advice, plans for reviews of manuscripts and, if the resources can be found, a later meeting to discuss the research results. Depending on the topic and the extent to which the course emphasized training, a course of this kind could last for one or for two weeks.

Each summer session should include at least one course, perhaps a short seminar, oriented around a substantive research topic. The idea would be to gather researchers to review work in an area and discuss research priorities. This should be coupled with a broad examination of the available data sources relevant to the topic. Hopefully one would bring together a mix of junior and more senior researchers to stimulate

research in the chosen area and the organizers should develop a strategy to encourage ongoing communication among the researchers.

On a revolving basis, the summer program should offer courses on a variety of topics. An attractive idea would be a seminar on policy research with quantitative databases. Researchers from government departments, as well as policy organizations, could be brought in to describe their own work and current policy and research concerns in their areas. A short course on data concepts should also be considered. The idea would be to focus on the way that Statistics Canada, and survey researchers in general, conceptualize and measure the parameters of key social variables. Some obvious candidates include race and ethnicity, labour force participation, work experience and perceptions of the quality of the social environment.

Because there appears to be no Canadian graduate program with advanced training in survey design, consideration should be given to including one in the summer program. Training in data analysis should produce researchers who are aware of the limitations of their data, but is not a substitute for knowledge of the extensive literature on survey design. Ideally, a course would begin with a review of research on questionnaire design and strategies for pre-testing, evaluating and revising new surveys, then turn into a workshop where the participants would write, test and revise a new questionnaire. The second stage of the course would ideally be conducted in co-operation with Statistics Canada or a university-based survey research organization.

Appendix C

RESEARCH DATA CENTRES

To facilitate access to confidential micro-data for research purposes, researchers would become "deemed employees" of Statistics Canada. There are numerous legal and organizational issues that must be addressed under this scenario, and this appendix outlines one approach. One issue is physical proximity to the data. Access is difficult for many if the data are available only in Ottawa. Hence, it is proposed that this approach be developed within a framework of Research Data Centres.

1. WHAT IS A RESEARCH DATA CENTRE?

A Research Data Centre would be a physical location that would have a secure environment capable of protecting confidential micro-data files, and would be an extension of Statistics Canada. It would have an affiliated research program administered by SSHRC and Statistics Canada. Researchers sworn in under the Statistics Act would have access to confidential micro-data files maintained at the centres, thus allowing research to proceed. Researchers would be held accountable for the protection of confidentiality in exactly the same way as Statistics Canada employees are currently held accountable.

2. THE MAJOR ISSUES ASSOCIATED WITH DEVELOPING A RESEARCH DATA CENTRE PROGRAM

The protection of confidentiality. This is of central importance, since confidentiality is the cornerstone of the statistical system. Any access to data by researchers who are not regular employees of Statistics Canada must be done in accordance with the Statistics Act.

Ensuring that the work and the researchers fall under the Statistics Act. In order to ensure confidentiality, the researchers must be sworn in under the Statistics Act, and be subject to the conditions of the Act in the same manner as a regular Statistics Canada employee. The Act lays out the conditions under which people can be given access to confidential data, and hence these conditions must be incorporated into any program of the Research Data Centres.

Ensuring that the Centres will succeed in their objective. The centres must be structured in such a way as to attract top quality researchers and their students to ensure that the goals outlined earlier in the report are met.

Ensuring an appropriate organizational structure. There are issues regarding the manner in which such data centres are organized and managed and the composition of the approval and evaluation committee for the research proposals. It is proposed that SSHRC and Statistics Canada jointly administer the affiliated research program, and that Statistics Canada manages the facility, as it is an extension of that agency. Operationally, the selection and vetting for the research program would be conducted by committees consisting of prominent researchers, in compliance with the requirements of the Statistics Act and confidentiality.

Ensuring appropriate physical safeguards for the data. To both meet confidentiality requirements and to openly demonstrate that they are being met, adequate physical security obviously must be in place.

These issues are addressed in the following outline.

3. THE GOALS OF THE RESEARCH DATA CENTRE PROGRAM

The program would have two goals:

- To promote quantitative research by academics, researchers in government agencies, research institutes or elsewhere in the public sector using Statistics Canada micro-data files, particularly household micro-data files.
- To improve the statistical programs of Statistics Canada through feedback from researchers using the micro-data, and through the research and papers produced by the researchers.

4. ELEMENTS OF THE RESEARCH DATA CENTRE PROGRAM

Following are the essential elements of a research data centre program:

- A secure, but user-friendly computer environment in which confidential micro-data could be stored would be required. Security would be at the standard maintained by Statistics Canada.
- A selection, approval and vetting process for the research based on confidential micro-data. A committee consisting of senior researchers (from academia and elsewhere) in a number of fields, as well as SSHRC and Statistics Canada officials would be created to administer the selection and vetting process. By and large, senior academics would determine which projects would be approved. There would be one peer review panel for each major area of research, such as economics, sociology, health, education, and statistical methodology¹. Academics or researchers at research institutes, government agencies, or other research organizations could submit proposals.
- There would be at least one Statistics Canada employee on-site at the Data Centre, to manage the site, provide support and to oversee confidentiality issues.
- Only persons sworn in under the Statistics Act through the Research Data Centre Program would have access to data at the Centres. Researchers would sign a contract that dealt with their obligations under the Act².
- The Centres would have to be self-financing, with funding coming from the institutions running the Centre, granting councils such as SSHRC, or through the researchers.

5. OPERATING THE CENTRES UNDER THE STATISTICS ACT

Researchers with access to confidential data must be sworn in under the Statistics Act. Aside from regular Statistics Canada employees, the Statistics Act restricts access to confidential data to people who are "retained under contract to perform special services for the Minister (i.e. Statistics Canada)." These people become deemed employees of Statistics Canada. That agency retains researchers to conduct work when it does not itself have the resources to do all the work required. Statistics Canada clearly does not have sufficient resources to produce the analytical work needed to exploit the many new (and older) data sets. Such exploitation is needed to provide analytical insights of value in public policy development and debate, and in the promotion of basic research. Much of the capacity to take advantage of the rich data resources resides in the academic community and in other research organizations (in other federal agencies, think tanks, provincial governments, etc.).

In this environment, and under the current Act, in order to be sworn in under the Act the research done would be similar to that which Statistics Canada itself would normally conduct (if it had the resources). While this may sound restrictive, in fact Statistics Canada carries out a wide range of research; consequently, this is not likely to be an issue. In the selection process, the merits of the proposals would be paramount. This selection and review committee would oversee the review process after the completion of the project. Statistics Canada would obviously have some input to the process, since under the Statistics Act the work constitutes special services for the Minister.

Seeking approval from respondents to share the data with data centre users. There is an alternative means by which confidential data can be shared with users who are not covered by the Statistics Act. The Act includes a provision for the sharing of data with users. If the respondent's permission is sought at the time of data collection, confidential micro-data may be shared with selected users. Data sharing would be possible with SSHRC or an "Institute" created by that agency. Respondents would be informed about the nature of the institute (or SSHRC), and permission sought to share the data. Based on past experience, it

¹ Preference might be given to empirical research that has policy relevance. To allow the system to be responsive to rapidly emerging policy questions, a two-tier approval process should be considered. One tier would be designed to render quick decisions for short-term feasibility studies with only minimal funding, and another directed at longer-term projects with more substantial resource implications.

² In order to promote the training of a new generation of researchers with expertise in the use of these data sets, and in order to facilitate the research itself, this should include research assistants, especially graduate students, employed by the academic researcher or under his/her supervision in dissertation research.

is likely that most respondents would agree. Under such a scenario, Statistics Canada would then provide the raw micro-data to the Institute (or SSHRC), which would of course agree under a contract arrangement to maintain confidentiality.

This approach has the advantage of being straightforward. The selection and vetting process would then be entirely in the hands of the Institute; Statistics Canada would play little role other than setting up the conditions to ensure data confidentiality. The shortcoming of this approach is that it can only be applied to data collected in the future, and to some data sets. The data sets that have been created to date could not fall under such an arrangement. For that reason, this approach should be seriously considered for the future, but will not solve the current issues. Thus, there is the possibility of forming an "Institute" with which data sharing could take place. The propensity of respondents to share the confidential data with such an institute (or SSHRC) for research purposes could be tested. While this approach would have little benefit in the short run, in the longer run it is a potentially excellent solution to sharing selective data sets.

Handling the research paper upon completion of the project. Researchers may want to think in terms of two papers following completion of the project. A research report consisting of the quantitative analysis and an interpretation of the findings would be deposited with the Selection and Review Committee of the Data Research Centres. This product would become part of a research paper series³, and would be reviewed in a normal academic peer review manner. The Committee (or its named designate) would run this process. The product would also undergo an "institutional" review by Statistics Canada. This is simply to determine if the work falls within the mandate of the agency for the purposes of the research paper series. Work conducted for the research paper series could be (and often should be) policy relevant but could not contain direct policy recommendations. Hence, the institutional vetting by Statistics Canada is conducted primarily to ensure that there are no direct policy recommendations in the paper⁴. After the academic and institutional reviews, authors will make the necessary changes.

The researchers would, of course, be free to publish the research paper, or an extended version of the research paper with policy comment and other additions they see fit, in any academic journal or any other venue. In short, beyond the initial quantitatively-oriented paper that is submitted to the research paper series, the researcher would be free to produce any other version of the paper, and submit it for publication in any forum. In the event that the review of the paper by the Selection/Evaluation Committee and Statistics Canada leads to a decision not to publish the paper in their research paper series, the author will be able to submit it for publication elsewhere⁵.

³ The research paper series developed for the publication of the initial quantitative analysis would likely be best maintained in one central location. Managing the research paper series centrally ensures that there is some consistency in the way the papers are handled. There may be resource considerations for Statistics Canada and the Committee. Vetting the research papers to ensure that the work is properly conducted may at times be time consuming. Such vetting will be largely done through the academic refereeing process. At times, however, work may be required to validate the use of the data. It is important that the person responsible for maintaining the series understands research and the research world, and has the resources necessary to prevent a backlog of papers.

⁴ Direct policy recommendations or comments refer to direct evaluation, criticism or advocacy of existing or proposed government programs. This does not exclude research on policy relevant topics, which is in fact encouraged. These restrictions are placed on Statistics Canada output (and the joint research paper series) to protect its neutrality and objectivity.

⁵ Normally the copyright on work that has been done under contract as a special service for the Minister (as required by the Act) would be vested with Statistics Canada. However, in the contract struck before the work begins, it is proposed that Statistics Canada agree to vest the copyright with the researcher. Statistics Canada would retain the right to vet all publications stemming from the project for confidentiality and data reliability. In the same contract Statistics Canada would retain the right to reproduce the paper if it chose to do so.

6. WHERE THE CENTRES MIGHT EXIST

The centres may be affiliated with a university, research institute, or research network. They could also be located at Statistics Canada regional offices, and for reasons outlined later, it is proposed that regional offices be employed initially. Obviously a physical location with a secure environment is needed. If located outside of Statistics Canada, competitions could be held to determine where such centres might be placed. One would want to start with a very small number, as there would be a substantial resource impact on Statistics Canada through the servicing of such centres. It seems likely that having such a centre would allow the institute serving as a home base to develop a very strong empirical research capacity in potentially a number of disciplines. Very qualified empirically oriented researchers would be attracted to an institute with a Data Centre, allowing the university or institute to develop a strong program.⁶

7. STARTING SMALL

This is an ambitious program. A pilot approach may be appropriate. It is proposed that the data sets made available to the centres initially include the new longitudinal surveys: the Survey of Labour and Income Dynamics (and the Labour Market Activity Survey, its predecessor), the National Longitudinal Survey of Children and Youth, the National Population Health Survey, the Workplace and Employee Survey, and related social statistics data sources required to conduct the research. It is here that the need for direct access to the micro-data is the greatest. It is also proposed that regional offices be the initial sites for the pilot project. This would allow the development of the centres to proceed incrementally. Nonetheless, access to the confidential micro-data would be improved tremendously, and some fairness regarding the geographical location of the centres would be introduced, as regional offices are located across the country. If housed in Statistics Canada, the initial centres may be less costly to run, as some of the infrastructure already exists. After the pilot project has been in place for some time (perhaps after two years), it is proposed that the program be reviewed, and subject to this review, a competition for Research Data Centres at Universities or other research organizations be held.

8. THE FUNDING

Similar centres operated by the U.S. Census Bureau and the U.S. National Science Foundation have an annual budget of U.S. \$250,000 per location. There are at least two alternative funding approaches. The first is similar to the way in which the Data Liberation Initiative (DLI) is financed. Under this scenario, the Centres would be entirely block funded. Such block funding would be provided by SSHRC, the university or institute at which the centre is located (after the pilot phase), and other organizations whose researchers use the centre, such as government departments, research institutes, etc. Since it anticipated that primarily academic researchers would use the centres, the majority of the block fund would be SSHRC-based. There would be no direct cost to the researchers using the centres. The selection process run by the selection and review committee would regulate access to the centres. A variant of this approach would be to have much of the annual cost covered by block funding, and the remainder covered by fees paid by researchers using the centre. University-based researchers whose projects were approved by the committee would receive a SSHRC grant (distributed through the review committee) to cover this cost. Researchers from other organizations would pay similar fees. It is proposed that one of these funding approaches be implemented.

⁶ Certainly the NBER (National Bureau of Economic Research) in the U.S., one of the first institutes to have a Census Bureau Data Centre, has developed a very strong empirically based research program, presumably in part due to the superior data access available to researchers affiliated with the NBER.

9. Conclusion

This proposal urges the consideration of Research Data Centres as one solution to the issue of access to micro-data for researchers. Initially this would involve access to household survey data, particularly the new longitudinal household surveys. Business (establishment or company) surveys present unique confidentiality issues that may prevent them from being included in this endeavor, at least initially. The Centres could be supplemented by a remote access capability developed within Statistics Canada. The latter would be useful for researchers who are not geographically capable of using a Data Centre, or who for other reasons cannot or wish not to use the Data Centres. Smaller projects, for example, may be better dealt with through remote access.

Appendix D

THE DEVELOPMENT OF A REMOTE ACCESS CAPABILITY AT STATISTICS CANADA

A Remote Access Capability should provide researchers with tools allowing them to specify statistical procedures and to have these procedures applied to confidential data by Statistics Canada staff. The results would be screened by Statistics Canada for confidentiality and returned to the researchers.

Among the tools available to the researcher would be the public use micro-data file (where available) and a permuted file which closely mirrors the confidential file in structure and detail, but with only a limited amount of real data. The researchers would specify their program or desired run as an SPSS/SAS job, using the permuted file, and then submit it to Statistics Canada via the Internet. The permuted files should contain sufficient detail to provide a basis for the testing and debugging of programs. Statistics Canada would not play a role in such debugging; program errors would be returned to the researcher. The main task of Statistics Canada employees would be to run the jobs against the master file and vet the output for confidentiality. Programs that led to confidentiality problems would have to be modified by the researcher.

The main advantages of remote access are its availability to researchers regardless of their physical location, and the simplification of the interface between them and Statistics Canada. Its disadvantages are its costs (when compared to the direct use of publicly available data files) and the possible delays incurred as the results are vetted. An informal working group has been struck within Statistics Canada to develop an implementation strategy that would minimize these disadvantages. For the moment, it goes as follows: the files would be produced by the subject area responsible for the survey, with the assistance of methodology staff skilled in the generation of such files. They would be distributed to researchers via the Data Liberation Initiative (DLI)/ FTP site; DLI contacts at universities could help to facilitate access to the files as they do with the current DLI files. Statistics Canada would establish a central group, building on the present DLI team, to run the programs and screen the outputs for confidentiality. The results would be returned to the researchers using FTP.

The main challenges about this approach have to do with the range of permuted files that can be produced and made available, and with the effectiveness with which this can be done, with the effort required to screen the output for confidentiality, and with the turn-around time. Many of these questions will be answered only after a suitable test of the strategy. The latter has been implemented for the National Population Health Survey (NPHS) and it is anticipated that there will be four files available for testing in a remote access environment next year. This should provide researchers and Statistics Canada with enough information to decide how remote access can be used, in conjunction with other approaches, to improve access to data.

Appendix E

A MODEL FOR RESEARCH FORUMS IN SOCIAL STATISTICS: THE CANADIAN EMPLOYMENT RESEARCH FORUM

CERF is a non-profit corporation, whose primary goal is to increase productive interaction between researchers studying employment issues in the academic and government sectors, as well as policymakers themselves, and to increase both the volume and quality of policy-relevant research in this area. CERF is directed by a rotating volunteer board, at least one third of the directors come from the academic community and from the government service. CERF receives core funding from Human Resources Development Canada, but it also raises considerable additional funds for its conferences from granting agencies, foundations and stake-holding government departments.

In the last eight years, CERF has organized a series of conferences (close to twenty, so far) on a wide variety of research topics. As a result, CERF has dramatically improved the access of policy makers to the technical expertise of academic researchers, at very low cost. CERF also sparked a dramatic increase in research using Canadian data. This was no accident: CERF conference organizers encouraged Canadian research, though they recognized the benefit of comparisons with the U.S. and other nations. Through contacts in government, CERF also facilitated access to confidential data. A list of conferences organized by CERF follows.

- Founding CERF Conference. Kingston, Ontario, May 31, 1991.
- Policy Research in Training, Unemployment and Income Support, and Immigration. Aylmer, Quebec, March 6-7, 1992.
- Labour Markets in the Last Two Recessions: A Comparative Perspective (with Statistics Canada). Ottawa, Ontario, March 5, 1993.
- Income Support Programs and Policies (with Health and Welfare Canada). Ottawa, Ontario, September 24, 1993.
- Immigration and the Labour Market. Hull, Quebec, March 11, 1994.
- Labour Markets and Income Support: Focus on British Columbia (with B.C. Ministries of Social Services and Skills Training). Vancouver, B.C., March 25, 1994.
- Workshops in conjunction with the Learned Societies Annual Conference. Calgary, Alberta, June 13, 1994.
- Youth Labour Market Adjustment (with BC Ministry of Education and UBC), Vancouver, B.C., June 25, 1994.
- International Conference on the Evaluation of Unemployment Insurance. Ottawa, Ontario, October 14-15, 1994.
- Policy Responses to Displaced Workers (with the Université du Québec, Montréal and Ekos Research Associates). Montreal, Quebec, December 2-3, 1994.
- Retooling the Employed Workforce: Focus on New Brunswick (with HRDC, NB Region and UNB). Fredericton, N.B., March 31 - April 1, 1995.
- Environmental Policies and Labour Markets (with HRDC Innovations). Ottawa, Ontario, Sept. 15-16, 1995.
- The Canada - U.S. Unemployment Rate Gap (with Centre for the Study of Living Standards). Ottawa, Ontario, Feb. 9-10, 1996.
- Changes in Working Time in Canada and the United States (with Upjohn Institute, Statistics Canada, HRDC). Ottawa, Ontario, June 13-15, 1996.
- Social Experiments, Evaluation and Employment and Social Policy. 1996.
- Immigration, the Labour Market, and the Economy. Richmond, BC, October 17-18, 1997.
- Participation Rates and Employment Ratios. Ottawa, Ontario, December 17, 1997.
- Labour Market Transitions and Income Dynamics (co-sponsored by Statistics Canada and HRDC). In conjunction with the CEA meetings in Ottawa, Ontario, May 28-29, 1998.
- Self Employment Conference (with the Canadian International Labour Network and the OECD). Burlington, Ontario, September 24-26, 1998.

More information on CERF is available at its website: <http://cerf.mcmaster.ca>

Appendix F

MEMBERSHIP of THE JOINT WORKING GROUP

Paul Bernard, Chair
Département de sociologie
Université de Montréal

Betty Havens
Community Health Sciences
University of Manitoba

Peter Kuhn
Department of Economics
McMaster University

Céline Le Bourdais
INRS – Urbanisation

Douglas A. Norris, Director
Housing, Family and Social Statistics Division
Statistics Canada

Michael Ornstein
Institute for Social Research
York University

Garnett Picot, Director
Business and Labour Market Analysis
Statistics Canada

J. Douglas Willms
Atlantic Centre for Policy Research in Education
University of New Brunswick

Martin Wilk, Former Chief Statistician
Statistics Canada

Assisted by:

Hélène Régnier, Policy Analyst
SSHRC