CATALOGUE 11-005 VOLUME 4 / NUMBER 4 / JANUARY 1972

STATISTICAL OBSERVER

CONTENTS

STATISTICAL OBSERVER

VOLUME 4 / NUMBER 4 / JANUARY 1972

The Statistical Observer is designed to contribute toward informing economists, statisticians and related professionals throughout Canada about selected statistical and research developments undertaken in Statistics Canada, in other federal departments and agencies, in provincial departments, in universities and in business and independent research organizations.

Suggestions, contributions of articles for publication, and requests for subscriptions should be addressed to The Editor, Statistical Observer, Statistics Use and Information Services, Statistics Canada, Ottawa K1A 0T6, Canada.

ISSUED, FREE OF CHARGE, BY STATISTICS USE AND INFORMATION SERVICES, STATISTICS CANADA, OTTAWA.

CATALOGUE 11-005

Published by authority of the Minister of Industry, Trade and Commerce

0804-506

The Computer and Government Statistics

Technological barriers to the effective realization of many of the overly enthusiastic pronouncements of the sixties regarding the use of the computer are being removed, and the cost of hardware is actually declining. Yet one can sense in the literature on computer use a tendency toward greater caution. This tendency is presumably the result of the failure of many automation projects to fully realize their promised goals, and to complete their implementation within anything like their scheduled times. The present paper *(2)* is an attempt to present some views on automation in the light of the recent experience in Statistics Canada.

We shall concentrate here on problems associated with the automation in a statistical office. In so doing it is our intention to promote the cause of automation rather than to inhibit it. By identifying problems and facing them squarely we should be able to cope with them more effectively.

Some Reflections on the Goals of a Statistical Office and Automation

Clearly the goals of automation are derived from the goals of the statistical office which, in turn are derived from the goals of society. It is appropriate then to start with the question why should a statistical office aim to automate *(3)* its operations?

There are, of course, the usual considerations of increasing output per dollar spent and increasing output without a substantial growth in manpower, thus making the statistical office more efficient and effective. However, considerations of effectiveness become much more prominent when we consider the basic goals of society which should be reflected in the goals and programs of the statistical office. The goals of society are seldom defined unambiguously. The more recently formulated social goals have so far been articulated in most societies in very broad and vague terms, such as the elimination of poverty, abatement of pollution, a more egalitarian society in terms of income and asset distribution, etc. Precise definitions of these notions and measurable concepts are still to be worked out and a great deal of experimentation is involved, by users and producers of statistics alike. Thus, the statistical agency finds itself facing a multiplicity of goals, not fully defined in advance. For achieving these goals, the traditional cycle of statistical operations consisting of collection, processing and publication is becoming too inflexible. What is required, in addition, is the storage of the most detailed data and a capacity for retrieval of an almost infinite variety of unanticipated cross-classifications and aggregations, consistent with confidentiality requirements.

There is also a growing requirement to be able to interrelate data collected in a variety of surveys or through other sources. In fact, policies and programs are frequently interrelated so that the formulation of any single one must take into account its effect on the others. Broad approaches to the analysis of problems requires a great deal of detailed data and the most flexible retrieval capability facilitating their joint and interrelated use. The more widespread use of economic and social models based on a large number of equations and the development of simulation models designed to study the behaviour of persons, business and governments, as well as society as a whole, under alternative assumptions, clearly require massive data storage and quick and flexible retrieval capability, as well as an unprecedented level of data cleanliness (editing).

The preceding suggests the goals the statistical office should try to reach through automation, but the attainment of these goals is enormously difficult so that they are probably best regarded for the moment as merely indicating the *direction* in which the statistical office should strive to go. We can no more than hint here at the prerequisites for achieving the goals and at the immensely complex obstacles in the way, before turning to more *immediate* problems which must be coped with to carry forward successfully a practical program of automation.

First, as already suggested, the statistical office must develop facilities for storing the data collected in highly disaggregated form, (to be referred to as "microdata sets") and for retrieving quickly a large variety of unanticipated as well as anticipated tabulations. The storage of microdata has major implications for processing: the microdata must be particularly well edited, adjusted for non-response, weighted in the case of sampling, etc., and these operations must be carried out in relation to individual records. Thus statistical standards related to concepts, classification, edits and imputation must be much more rigorous than formerly when the statistical end products were, by comparison, much fewer in number, more highly aggregated and largely preconceived.

Second, statistical standards related to collection, (coverage, field work, etc.) must be tighter than previously to ensure that the large variety of detailed tabulations and cross-classifications will have predictable (and acceptable) margins of error. This means an increasing responsibility to follow up the non-respondents, probably involving a greater role for the field function, greater integration of follow-up procedures for different surveys which, in turn, necessitates careful scheduling and planning. Put differently, the integrity of the data must increasingly be assured during its collection and processing at the level of the individual observations since surveys can no longer be designed with a view to minimizing the error of a few predetermined aggregates only.

Third, the more data are accumulated in the statistical office, the more important it is for the statistical office to carry out analyses of the data in order to identify the relationships existing between the data derived from different surveys and other sources, with a view to providing users with information about the character, limitations and potentials of the data. The more data is available, the more important it is to standardize concepts, study and weed out inconsistencies, structure the data base in a coherent manner and, in general, to carry forward what is known as the function of statistical integration. In sum, the statistical office should strive to provide an analytical infor-

mation service, rather than just a data collection facility – a very challenging task.

The last problem to be mentioned here is that of privacy and confidentiality. Previous practices of publishing only predetermined aggregates lent themselves more easily to confidentiality checks, in contrast to the retrieval flexibility mentioned above. The latter carries with it dangers, which must be overcome, of inadventently disclosing confidential information. Moreover, the very process of storing individual returns in machine-readable form and building up facilities to interrelate them creates fears in the public mind in regard to their privacy. These fears must not be disregarded on the grounds that they are unfounded because statistical offices merely require individual observations as raw material to produce aggregations which remove identifiable events.

Some of these points are discussed later and we now turn to the first of our immediate practical problems with regard to automation, namely, data problems.

Data Problems

At the heart of most automated survey processing systems is a subsystem to accomplish the editing of individual returns, their correction and the imputation for non-respondents. The proper use of the edit and imputation subsystem in an automated survey is vital to overcoming many of the data problems involved in using flexible retrieval systems. However, this subsystem is typically the most complex of the automated survey processing system and great care must be taken that its implementation does not get out of control (in terms of time, resources, testing the results, etc.).

It appears to us that automatic editing and imputation must have a dual role in the processing of systems. The first role is the most visible one, namely the identification of inconsistencies in the reported data and their elimination, the identification of non-respondents and the creation of some imputed data for them. The second role is to provide printouts indicating, in summary form, the impact of the changes in the data due to computer processing. The statisticians must be able to assess the impact of the changes in the data arising from such processing and if necessary, intervene to make manual corrections.

The most important intervention probably relates to non-respondents. It is unrealistic to assume 100 per cent response rates even though we aim to approximate it. Computer editing and imputation and the resulting summary measures can focus the follow-up effort in the areas where it can have the most important impact.

In this connection, it is important to emphasize that the storage of microdata and the availability of flexible retrieval systems should have an impact on what are considered to be important or unimportant cases of non-response. Traditionally when we were aiming to produce only a few major aggregates from a survey, we identified as the most important non-respondents the large units; or, in the case of repetitive surveys, units which had a volatile pattern of reporting from survey to survey thus making imputation on the basis of historical data difficult. This strategy needs to be reassessed in the era of microdata set storage. Some of the unanticipated retrieval requirements may well relate to small survey units for which we traditionally tended to allow a higher non-response rate. At the very least, in addition to following up the "important" non-respondents, we should also follow up a sample of the remainder in order to evaluate the impact of imputations on all types of non-respondents.

Increased flexibility of retrieval may expose data weaknesses not only within individual surveys, but also between surveys. One of the objectives of automation of individual surveys is the creation of a data base permitting easy retrieval of information from the files resulting from the particular survey. In a sense, therefore, we are in the process of developing a multitude of small data banks whose subject matter scopes overlap. As more such data banks are developed, they can create incompatibility problems of immense complexity. Integration tools are urgently needed to overcome these problems. One important integration tool is a central register including the universe of units included in individual surveys. Such a register can facilitate an unambiguous definition of the scope of different surveys, it can provide a high quality frame for the selection of samples, it can provide a tool for the consistent classification of identical units in different surveys, and it can facilitate the comparison of microdata collected in different surveys. Statistics Canada is in the process of establishing a central register covering all business (and other, for example, institutional) reporting units, taking into account the many complexities in the identification of respondents in different business surveys.

At least as important as a uniform frame for related surveys is the problem of uniform concepts. What is needed with increasing urgency is a data element dictionary identifying the statistical concepts underlying each data element (for example, "hours worked", "total retail sales", etc.) which can serve as a standard convention enabling the unique referencing of identical data elements with identical terms and different data with different terms. Such a data element dictionary would also serve as the foundation for an overall system of file descriptions of all machine-readable files. Clearly, such description system is a necessary part of the strategy of flexible retrieval and other statistical manipulations for we must be able to reference identical data by the same terms for the convenience of users.

One of the most important data problems facing the statistical office is to draw users' attention to the degree of reliability of the statistics released. We must develop general models incorporating the contribution to the total mean squared error of sampling error, reporting errors, as well as conceptual, classification and processing errors. Although it might not be possible to estimate the mean squared error of each individual retrieved aggregate, it should be estimated for a sufficient

variety of such aggregates to enable the development of general models of the mean squared error which could provide guidance to users of the order of magnitude of the errors associated with the data.

Statistics Canada has developed a very flexible retrieval system which is being implemented in connection with the 1971 Population Census. We are in the process of developing generalized tables which will be provided to any user who obtains census data and which will show for each of several broad classes of estimates the size of the mean squared error as a function of the size of the estimate. Similarly, we have developed general tables for the sampling error of Labour Force Survey data. These efforts must be considered, however, only as the very beginning of a program to develop tools to enable users to identify and appraise errors.

An important part of controlling data problems and of developing the necessary error models is the conduct of a rigorous program of *evaluation* of surveys. We are used to computing sampling errors and, at Statistics Canada, the formulae used traditionally for the computation of sampling errors incorporate the impact of the random reporting and processing errors. Such formulae, however, do not measure the bias which may be due to any of a number of possible sources. Starting with the 1961 Census, we began to conduct, as part of our quinquennial population censuses, extensive evaluation programs aimed at measuring the different components of the mean square error. These evaluation studies had far-reaching impacts on the methodology employed in the 1971 Census. Several of our other surveys incorporate control features which provide as a by-product some evaluation type information. However, we have a very long distance to travel before we can claim to understand clearly the sources of errors in surveys and the complex forces which generate them.

Project Implementation Problems

Although data problems are probably the most fundamental problems underlying the automation of statistical processes, the most immediate and frustrating problems relate to project implementation.

Several of our surveys now use machine-readable files for mailing out questionnaires, checking-in respondents, the identification of non-respondents and the corresponding follow-up actions, the editing and correction of the returns and imputations for non-respondents, tabulations and variance tabulations. We are in the process of automating a comprehensive central register of business units. We have automated the data processing of the 1971 Census, including the editing, correction, sample weighting, the production of predetermined tabulations, as well as the creation of a general retrieval system to permit the retrieval of any cross-tabulation for any area based on the census data. We also have established a general time series data bank capable of storing tens of thousands of time series and retrieving any number of them for printout or subsequent manipulation. However, with a few notable exceptions, our accomplishments to date have been accompanied by considerable frustration and delay. We have made a conscious effort to learn from our own experience but there are still many lessons ahead of us.

In a paper *(2)* presented in London in 1969, we articulated five important lessons we learned from our previous automation experience. Although these lessons need to be supplemented, in the light of our subsequent experience, as far as they go they are as valid today as they were two years ago. These five points, in summary form, follow.

1) We have emphasized the importance of developing an overall design for the automation projects identifying its individual components and specifying in detail at least some of them before programming begins.
2) The ultimate system implied by the overall design should be split into fairly self-contained modules. The modules should be small enough for the implementation to be carried out in a reasonably short time by *average* programmers.
3) Specifications for the corresponding computer systems have to be developed in collaboration with experts in survey methodology, computer systems, and subject matter.
4) It was stressed that individual responsibilities must be assigned within a working team and that a project manager be designated.
5) Finally, the vital importance of effective communication between members of a project team was emphasized.

During the last two years, although we were trying to live by the prescriptions outlined above, we nevertheless encountered some difficult problems of implementation.

One of these is that the stress on the necessity for an overall design and for complete and unambiguous specifications resulted in some unduly long gestation periods. Having emphasized that computer systems, once implemented, are difficult to change, having emphasized the need for complete and unambiguous specifications, we tended to inculcate a sense of finality in those who have the task of writing such specifications. Yet, in spite of this, we did not succeed in eliminating the need for changes to specifications after programming began or even after some of the subsystems were developed and tested. This was due to the fact that the ultimate systems we were aiming at could not be thought through in complete detail in advance of their implementation.

Although modular programming helps when changes to programs have to be made, it appears that we have to extend the concept of modularity from programming to implementation. In fact, what we may need is a phasing of implementation. By phased implementation, we mean one of two possible things: either the implementation of a subset only of the overall system, with additional subsystems added to it gradually as required; or the initial implementation of a skeletal version of the whole system, with provisions for its subsequent expansion and refine-

ment. The latter is the one we now favor. We want to emphasize that we are not advocating compromising the ultimate objectives. Rather, we advocate, where feasible, their attainment through several generations of gradually more complex systems.

The process of phased (or gradual) implementation just outlined has several advantages. First of all, it reduces the complexity of design testing and implementation since, *at any point in time*, one would be aiming at a more modest incremental goal. The achievement of the more modest goals provides a convenient check point for all concerned. The particular features of the new systems can be tested thoroughly in a real application (rather than using just test data). Shorter implementation periods also give us a better sense of achievement.

This point is very important. Personnel who are responsible for the operation of the survey, are faced at the time of automation with the task of learning all the complexities a major new system. In addition, they are required to monitor and test the data produced by the new system and compare them with the results of the old system. (This is necessary both as a system test and as part of the effort to monitor and control historical continuity of series.) Since this kind of testing of a new system cannot be accomplished using test data alone, these personnel are often faced with the parallel running of the old system and the new. There are seldom sufficient skilled resources to do so. The consequences of this, in terms of a long period of double work load, inadequate testing, acceptance on the basis of insufficient evidence, etc. are obvious enough. By contrast, in a phased implementation of a mail survey, for example, the mail-out and check-in can be automated and implemented after a relatively short period of testing. The implementation of automatic mail-out might immediately save some resources which can be available for the testing of subsequent subsystems. The next module to be implemented might, for example, be the tabulation module (leaving, for the time being, editing and data collection a manual operation). Finally a simple version of editing and imputation may be implemented which can be refined gradually to more sophisticated versions.

Thus, the first major advantage of phased implementation is that it reduces the impact of automation on the operating personnel (or, rather, it phases this impact over time in a more realistic fashion). A second advantage of phased implementation is psychological. The successful implementation of each new generation of the automated system and its acceptance for operational use provides a psychological boost to everyone concerned with the development.

The third advantage of phased implementation is that the implementation of each phase provides management with a check point at which a decision can be made whether additional features of the intended overall system should be proceeded with eventually, proceeded with immediately, or abandoned altogether.

There may seem to be an inconsistency between the concept of phased implementation presented here and our earlier views *(2)* that "the computer is more than hardware equivalent of a host of clerks. The application of computer processing is a fundamental parameter which must be taken into account in the design of surveys with the aim of rendering efficient the survey as a whole, given its objectives. In addition, it is our view that in order to derive the full benefits of computer processing, one should in general plan to go all the way: to comprehensive automation of all phases of the survey".

We must emphasize that modular implementation is no substitute for an overall system: it is a way of accomplishing it with as little pain as possible. The *outlines* of the overall system embodying the final automated survey must not be shortchanged. The modules that are selected for implementation must be consistent with the overall design. *Moreover, they have to be conceived in a general enough fashion to be capable of expansion and changes.* To the extent, however, that these modules would also be required to be compatible with the existing survey (or with some relatively small modifications of it) this may undoubtedly be the source of some conflicts. As additional modules are implemented, previous modules may have to be changed somewhat. We believe that so long as these changes are anticipated, their impact can be minimized. Even so, it is probable that if one were to compare the cost of implementing the ultimate system at once with the cost of implementing it in a gradual fashion, phased implementation might well be the more expensive of the alternatives. However, such a simplistic cost calculation would be based on the assumption that the systems are successfully implemented on schedule. This may not be the case. Thus, one should look at the incremental cost of phased implementation as insurance money. It may well be worth *planning* to spend, say 50 per cent more on the implementation of the ultimate system to ensure that we do not have to pay an *unplanned* 300 per cent more.

Fortunately, important tools are being developed to assist us with the problem of modular implementation. A series of generalized programs have been developed in Statistics Canada (and elsewhere) capable of coping with a particular phase of processing of a variety of surveys. For example, we have a generalized program to produce address labels for mail-out purposes; another generalized program for the comparison of two files, for example in preparation for historical editing; another generalized program for editing and data correction; and retrieval programs capable of producing tabulations (including weighted tabulations) from a number of files.

We believe that these generalized programs and their extensions will play an increasing role in the automation of our surveys. They have several major advantages over special custom-made systems. First, because they are generalized and widely used, they soon become reasonably well debugged. Second, because they are widely used, they must be well documented. Third, their application reduces the implementation

time in that it avoids the time-consuming process of program design, coding, testing and debugging. Fourth, and most important of all, they facilitate experimentation and make subsequent changes relatively easy.

This last point needs to be elaborated. When specifications are prepared for a complex custom-made program, the consideration of alternatives is by necessity a theoretical exercise. Because of the complexity of what is to be implemented, to assess and test its impact would require very large volumes of test data and almost impossible amounts of manual calculation. Thus, very often one finds out the full impact of such programs on the data only after they are implemented at a considerable cost and effort. The most fundamental advantage of generalized programs is that, because of their ease of implementation, they facilitate the testing of alternatives at a reasonable cost.

While generalized programs may not have precisely all the features that would be required for the implementation of the ultimate system, they may have an exceedingly important role to play in the scheme of modular implementation: they provide an easy means of implementing skeletal versions of the ultimate system which may subsequently be replaced, if necessary, with custom-made programs at a later date. In fact, as successive versions of generalized programs are developed, they may be capable of coping with very complex and sophisticated processing requirements as well.

It is essential that the anticipated benefits of automated surveys, as well as their cost and time of implementation, be estimated in advance and that these estimates be checked out after implementation in the form of post-implementation audits. This is an important feedback of the automation process enabling all concerned to gradually improve their ability to estimate both costs and benefits, as well as enabling the management of the statistical office to derive general policy lessons that can be applied to future projects.

Confidentiality Problems

It is a salient feature of statistical information that it always relates to a well-defined population rather than to a particular individual respondent. However, if a population of interest is sufficiently narrowly defined it may contain only one respondent. If this respondent can be identified on the basis of the statistics, their release would violate statistical confidentiality. Much more difficult is the problem of so-called residual disclosure. Residual disclosure occurs when a number of tabulations are released, none of which violates confidentiality in itself, but which together would enable a user to deduce information about a single identifiable respondent. Unlike direct disclosure, residual disclosure is notoriously difficult to detect even in the case of preconceived tabulations. Obviously, a policy of flexible retrievals exponentially increases the problems of checking for residual disclosure, since each new retrieval must be checked against all the previous retrievals.

One of the authors in a recent article *(4)* has developed a precise mathematical theory dealing with tests of residual disclosure. In all but the simplest situations, the practical implementation of this theory would involve prohibitive amounts of calculations. How can the overriding objective of protecting the confidentiality of statistical returns be reconciled with the need to use as extensively as possible the data that has been collected?

In connection with the general retrieval system which is being implemented on the 1971 Census data base, we have developed a general approach to the solution of the confidentiality problem. As far as direct disclosure is concerned, the retrieval system automatically checks each tabulation cell to ensure that it satisfies the predetermined requirements of statistical confidentiality. (This test is generally based on the requirement that each aggregate must relate to more than a predetermined minimum number of respondents.) To avoid the more complex problem of residual disclosure each *ad hoc* tabulation will be subjected to a small amount of random disturbance. This will involve a random reallocation of a small proportion of selected respondents within the table. The particular strategy of reallocation is so designed as to minimize the impact of this random disturbance on the accuracy of the data. This random disturbance will prevent users from manipulating the retrieved statistics in order to derive other statistics which were not tabulated.

Although, the strategy of random disturbances will increase somewhat the mean square error of the statistics, this may well be a necessary price for making the data available in the detail requested. The procedure can be automated and thus it can be fast and inexpensive. Even apart from this consideration, however, the strategy of random disturbances appears to be a "necessary evil" since a rigorous checking for residual disclosure is impossibly complex even with our modern computers. Statisticians are well used to making compromises between the often conflicting requirements of cost, reliability and timeliness. It now appears that a new dimension of the compromise must involve confidentiality and reliability.

Some Issues Confronting Management

In spite of what has been a somewhat cautious tone throughout this paper, a logical analysis of what has been said leads to the conclusion that the process of automation is, frustrations involved notwithstanding, desirable and essential. This reasoning starts with the fact that the extensive use of data *already collected* is the only conceivable way of satisfying the variety of information requirements of society. Thus, there is a compelling need for storing microdata and creating automated, flexible retrieval systems. Starting from this premise, and considering the strain that such extensive use puts on data reliability, the cleanliness of the stored microdata becomes an essential consideration. Thus we are inexorably led to the necessity of automated editing and correction at the microdata

level. We are also led to other considerations of statistical standards, such as the necessity for complete and up-to-date frames for surveys and censuses which, in the case of economic surveys, renders a central register of business units essential. An important source of updating such a register is from the surveys which are using it currently. Thus, we are led to the desirability and necessity for using the central register of business units either in the form of a central mailing list or in the form of a co-ordinating device. This, in turn, leads to the need for creating a machine-readable central register and machine-readable mailing lists for individual surveys, which would be connected by an automatic process of information flows. Thus it seems that, starting from the nature of information use, we are led to the full process of automating individual surveys and creating co-ordinating devices. One of the challenges of the management of statistical offices is to recognize this fundamental unavoidability of automation and take every necessary step to smooth and facilitate its broad implementation.

Another challenge is how far and how fast to go. We have no magic solutions to offer. Individual decisions will have to be made according to individual circumstances. However, a few important considerations, basically related to costs and benefits are sufficiently general to be applicable to almost any situation.

Short-term costs are very real and highly visible, although not necessarily accurately predicted at the beginning of development. However, once developed, the cost of effective data dissemination is relatively small in comparison with the costs involved in the collection of statistics. We are firmly convinced that the benefits of the two systems mentioned earlier in the paper (the general data retrieval system for the 1971 Census and the general time series data bank) will be far in excess of the costs. However, the precise estimation of these benefits in advance is exceedingly difficult. Given the long developmental time required for such systems, any advance estimation of their benefits would necessarily involve forecasting the information requirements of users some years in advance – a very difficult task. In the case of the 1971 Census information system, one can speculate that relatively few significant uses of the system would repay for its entire development cost. A single bridge that will not be built at the wrong place, the more precise delineation of one or two urban renewal areas, a single instance of a more precise identification of poverty areas which might receive economic aid from the government, a few household surveys that need not be taken because the data will be available from the Census through the system – any one of these applications could repay the development cost of the system.

The estimation of benefits to users is particularly difficult because users are a heterogeneous group both with respect to their requirements and their capability of using large amounts of information. It is relatively safe to forecast that user requirements will increase with respect to the amount of data required, the type of disaggregations, the kinds of manipulations, the format and medium of the retrieved data, and the desired retrieval response time. However, users have a wide variety of requirements and varying levels of computer sophistication. If the statistical office wants to keep in step with the users having the most advanced requirements, then it must be prepared to take greater risks in terms of development costs. Given the long development time required by automation, we believe that only a general policy commitment to attempt to *keep ahead* of the information requirements of users will succeed in ensuring that the statistical information service does not *fall seriously behind* the demands put on it.

Thus cost-benefit considerations of automation, while they might reaffirm the ultimate need for automation, do not necessarily provide generally applicable guidelines for determining how fast we should proceed with automation in comparison with, for example, the alternative of undertaking several new surveys. Here again, we believe that the development and utilization of generalized programs should have a high priority because of the broad facilitating nature of such programs. The automation in a one-by-one fashion of individual surveys in a large statistical office like Statistics Canada is a gigantic task. Custom-made programs which are parts of complex systems are difficult to change, and, with the need for change imposed by changes in technology, methodology, and user requirements, just updating earlier automated systems can keep occupied half to two-thirds of a large programming staff. Modular programming can diminish this ratio. However, we believe that only through extensive development and utilization of generalized programs will this ratio be significantly improved in favor of new developments, rather than changes to existing systems.

Footnotes

(1) This article is a summary of a paper by I.P. Fellegi, Director General, Methodology and Systems Branch, Statistics Canada and S.A. Goldberg, Assistant Chief Statistician of Canada, presented at the Conference on the Role of the Computer in Economic and Social Research in Latin America, at Cuernavaca, Mexico, October 25 - 29, 1971. The views expressed in the paper are the personal views of the authors and may not be shared by other officers of Statistics Canada.

(2) This paper draws upon and updates an earlier one presented at the London meeting of the International Statistical Institute (I.P. Fellegi and S.A. Goldberg, Some Aspects of the Impact of Computer on Official Statistics, Bulletin, ISI, Volume 43, 1969).

(3) As used in this paper, the word automation consists of more than mere conversion of existing operations to the computer. In this sense, automation of surveys means their complete redesign taking into account the full impact of the possibilities opened up by computer processing with the aim of rendering efficient the survey as a whole, given its objectives, not just its parts.

(4) I.P. Fellegi: On the Question of Statistical Confidentiality. To be published in the March, 1972 issue of the Journal of the American Statistical Association. Copies are available from the author on request.

Escalation of Industrial Contracts

The following paper, prepared by Mrs. C.M. Jones, Head of Capital Expenditures, Prices Division, contains suggestions on the use of price statistics in the escalation of industrial contracts.

One of the major uses of price index numbers is in the escalation or updating of agreements which have a life of many months or years. The practice arose because it was often found easier to achieve long-term agreements if both parties to the agreement were protected against changes in price levels during the life of the agreement. Participants in contract escalation have come to believe that some of the element of risk is removed if one can avoid having to predict price behaviour over long periods of time. Traditionally, parties to such contracts look to sources of published or official statistics to use in contract escalation.

Possibly the most famous example of escalation is the periodic updating of benefits under the Canada Pension Plan through a complex formula based on the Consumer Price Index. Some wage agreements are also adjusted on the basis of changes in the same index.

The Prices Division's role in contract escalation is one of obtaining a fairly complete statement of the purpose of the contract escalation from one or both participants and advising on the availability of price indexes in which the concepts and content appear relevant to the type of contract under consideration. Selection of the specific indexes to be used is the responsibility of the parties to the contract. When a selection has been made, the Division provides the indexes for the duration of the contract *(1)*. Many users obtain their indexes from the Division's major publication, *Prices and Price Indexes*, catalogue number 62-002, monthly. In other instances, the participants agree to use data produced by the bureau but not published in *Prices and Price Indexes*. In such cases, indexes are forwarded to one of the parties on a Statistics Canada letterhead and are thereby classed as official statistics. If the number of indexes to be provided is large or if an unusual amount of clerical resources is required to prepare the data, a charge may be made for this service.

Because indexes must be rebased or otherwise reorganized from time to time, parties to long-term escalation contracts are urged to inform the Prices Division of their index selections to ensure that it is possible to maintain comparable indexes for the life of the contract. This suggestion is of particular concern for parties to contracts who are using or considering the use of the General Wholesale Index for contract escalation. Users are also urged to write contracts which are not upset by index revisions.

Although responsibility for index selection and method of escalation rests with users, the Division is able to provide the following brief outline of the factors most users seem to keep in mind when selecting indexes for the escalation of industrial contracts.

1) The purpose of the escalation – For complex industrial goods such as pumps, turbo-generators and the like, a common purpose of contract escalation is to protect the manufacturer from changes in the prices of his material and labour inputs over which he usually has little control. The purchasers who are parties to such contracts tell us they can then better evaluate competing bids where the differing forecasts of future price change are omitted from the contracts bid.

2) The components in the contract to be subject to escalation – Most parties select important material inputs and direct shop labour. Production equipment used, overheads and profits seem to be escalated rarely.

3) The weighting of components – The method of deriving the weights, and ultimately, the weights to be assigned to the components to be escalated must be established. Weights give the proportional representation of the items to be escalated.

4) Index selection – Select the appropriate indexes to use in the escalation on the basis of:

(a) the contracting parties' history of price changes for the materials under consideration.

(b) an examination of a variety of the indexes most logically related to the commodities under consideration.

In practice, this means that the most closely related indexes are sought but, in most instances, users also examine more general indexes to which the commodity under question belongs. Thus, someone requiring an index for carbon steel castings would also look at price movement of other types of castings and the indexes for the castings industry and the gross-weighted iron and steel products industries.

The commodity content, the terms of sale and the internal weightings of all the indexes under consideration should be examined. For example, an index can be for the appropriate commodity but for an inappropriate market with the result that the commodity index shows different price behavior from that experienced by the manufacturer choosing escalators. Also, as most industrial price indexes are selected to measure average Canada price movement for sales by domestic manufacturers of a given commodity to a given class of customer, large and small volume purchasers or purchasers in small segments of the market may experience quite different purchase price experience.

For these reasons, most users not finding a specifically appropriate material index find it more reliable to use a related but more aggregative index to serve their purpose. The main reason is that more aggregative numbers usually move more smoothly than less aggregative indexes, and are usually less subject to sharp unexpected changes in movement. However, such price indexes reflect the average movement of many prices and may not be precisely representative of the prices experienced by an individual company. Nonetheless, many users of this type of index are content with an aggregative number which expresses

price changes for a generally related group of commodities, industry or group of industries. Such users intend to incorporate in their contract bid the forecasted impact of more specialized price changes which may not be suitably reflected in the movement of the selected price escalators.

In addition, users are urged to keep in mind one of the main characteristics of a price index. A price index is designed to reflect how much more or less it would cost in successive time periods to purchase an identical basket of goods and services. Changes in *costs* associated with changes in non-price elements of costs such as quantities and qualities purchased, will not be reflected in the movement of the price index. Indeed considerable efforts are expended to ensure that such effects are excluded from index movement. Non-price factors which relate to such events as change in models, changes of customers' terms of sale or quantity discounts, are excluded deliberately from price movement. For example, during a period when prices were unchanged, a particular purchaser may lose his quantity discount because he buys in smaller quantities. In this instance, the purchaser's costs have risen although a price index would show no change in level. Users of price indexes in contract escalation need to consider the implications of these conventions when designing the particular format of their contract escalation.

(c) Charting the indexes selected for a period of at least 30 months and considering the implications of historical price movement on the contract escalation, in relation to the producer's price experience.

For example, most manufacturers desire escalation because they have experienced rising input prices in the past. If the most appropriate index selected has a suitable trend but displays sharp variations about the trend, the parties to the contract might consider the most appropriate method of smoothing out the sharp short-term changes in price behavior.

(d) If none of the usually used indexes seems to move in an appropriate fashion, consider less traditional sources.

For example, average hourly earnings which are often used in escalation clauses are sometimes not appropriate for the price movement of a specific trade important to the contract escalation. In this instance, basic wage rates in union wage agreements are a possible data source. Union wage agreements are filed with a number of government agencies and, in our experience, the unions themselves have been most helpful in providing information about wage rate levels or changes in rates.

Users requiring escalation for imported commodities might consider the employment of statistics published by foreign statistical agencies, such as the United States Bureau of Labour Statistics, and the Department of Industry and Trade, the Department of Employment and the Central Statistical Office of the United Kingdom. Statistics Canada has on hand many of the publications of these agencies and they are accessible to users.

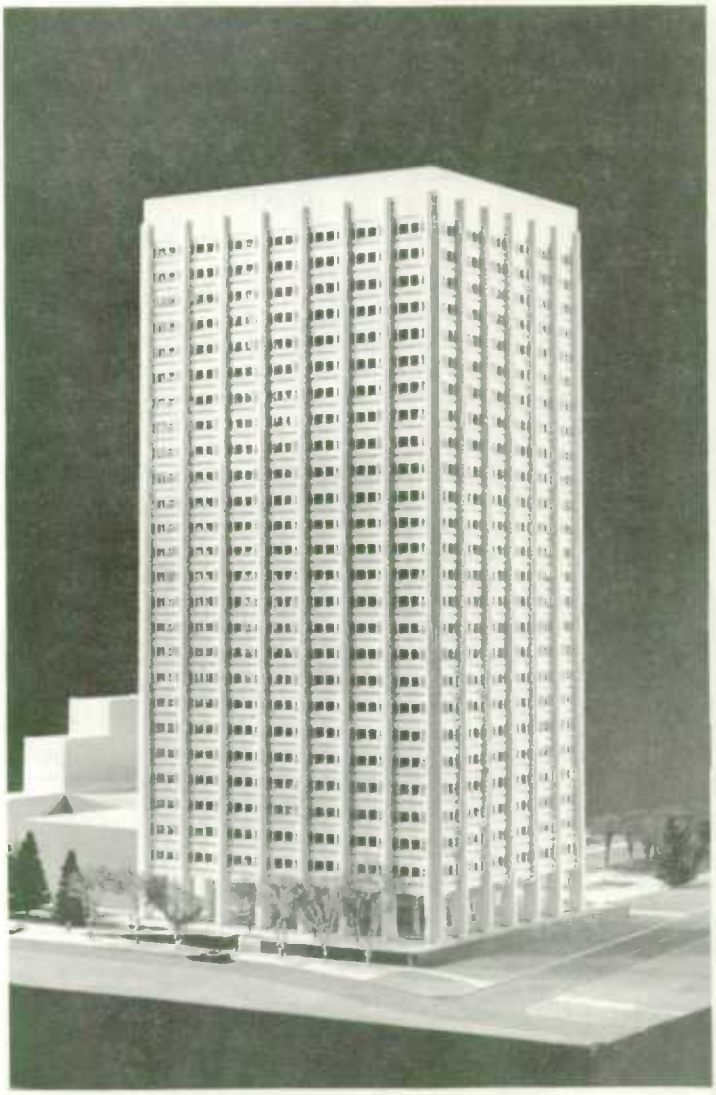
5) Design of escalation clauses – Once the index selection has been agreed to, the contracting parties should ensure that the contract is written in such a way that time base, title and other revisions to the indexes can be accommodated without invalidating the escalation portion of the contract.

Footnote

(1) This responsibility is subject to pre-contract discussion with the Prices Division.

STATISTICS CANADA TOWER – Construction is underway on a new office building for Statistics Canada. The new tower, which will provide 25 floors of office space above a mezzanine, ground floor and a basement, is expected to be completed in the spring of 1973.

The existing Statistics Canada building will be renovated and used to accomodate the administrative staff.



NEW PROJECTS

A New Look in International Travel Statistics

Statistics Canada, in co-operation with the Departments of National Revenue, and Manpower and Immigration, has launched a review of statistics on international travel. There are three phases of this program:

1) the implementation of an interim system of collecting numerical counts, by category of travellers, at entry points to Canada; and

2) The analysis of user requirements for international travel statistics, and the recommendation of systems of improved sample surveys to collect data on expenditures and other characteristics of international travellers.

The first phase has been completed and, on the basis of this analysis, an interim system (phase two) was introduced in January 1972. Because of the rapid growth in the numbers of international travellers and the need to facilitate the smooth flow of travellers at entry points without loss of important statistical data, the interim system is designed to improve reporting methods by reducing and standardizing forms, by implementing sampling techniques at some of the major entry points, and by centralizing the processing of the data at Statistics Canada headquarters.

The phase two system should make available more accurate and timely monthly data on the number of travellers by country of residence, by province and by port of entry. The information published, although similar to previous years, will be expanded to include the following data previously only available annually:

a) overseas residents categorized by entry direct to Canada and entry via the United States;

b) returning Canadian residents from overseas via the United States;

c) returning Canadian residents from the United States by private plane and boat; and

d) greater detail by port of entry.

A description of survey methods and definitions of terms used will be presented with the above data on traffic movements in the monthly publication *Travel Between Canada and Other Countries*, catalogue number 66-001. The first report will be available, in a bilingual form, early in 1972.

The third phase of the review began in the fall of 1971 with the listing and examination of user requirements. The results of this analysis should lead to possible alternative systems, the designing and planning of tests, the examination of trials and final recommendations. These steps will be carried out during 1972 with a more definite outlook expected for early 1973.

More information may be obtained from M. Valiquette, Chief, International Travel Section, Balance of Payments and Financial Flows Division, Economic Accounts Branch, Statistics Canada, Ottawa, K1A 0Z8.

Road Transport Review

A major review of road transport statistics has been undertaken by the Transportation and Public Utilities Division of Statistics Canada. Assessment teams have been established within the Division to critically examine each of the following surveys:

- roads and streets mileage and expenditure; international toll bridges, tunnels and ferries
- motor carriers, freight; moving and storage, household goods; warehousing
- trucking – origin and destination;
- motor vehicle traffic accidents; the motor vehicle, parts I to IV;
- passenger bus statistics; urban transit.

The assessment teams work under the general direction of an overall Project Management Committee. Each team is headed by a Project Manager, assisted by professional and other support staff who provide guidance and consultation in the course of the review of each survey. Each team is charged with the responsibility of assessing all aspects of each survey under review with a view to revamping the whole series of road transport statistics to more adequately reflect current conditions in the industry.

To date, the major emphasis of this review has been directed toward assessing each survey in terms of its usefulness to major users. All industry-oriented associations have been contacted and advised of the review and their co-operation solicited to assist in its conduct. The Ministry of Transport and the Canadian Transport Commission have been similarly advised, and representatives from both industry and government are actually working with the teams in each of the areas involved. This is considered a unique and highly effective method of ensuring their full co-operation and support.

Major respondents are also being contacted to ascertain changes in their record-keeping systems and practices, and to discuss reporting problems and collection procedures.

Although the review has been going on for only the past couple of months, already a number of changes and improvements have been recommended, such as publishing information in service bulletins to improve timeliness, revisions and additions to various series to improve their usefulness as well as a number of innovations and techniques designed to improve collection, processing and compilation procedures. The broad nature of the investigation is expected to result in additional recommendations of this kind as the review progresses.

Additional details concerning this review may be obtained from the Transportation and Public Utilities Division, Economic Statistics Branch, Statistics Canada, Ottawa, K1A 0T6.

Revision of Standard Commodity Classification

Canada's Standard Commodity Classification (S.C.C.) is undergoing its first major revision since 1959. The revision (more an updating than a change in basic concepts) provides new commodity classes and variations in old ones to reflect advances in technological developments and changes in the kinds of analytical data needed.

Publication of the first volume of the S.C.C. manual, entitled *Standard Commodity Classification Manual, Volume I – The Classification,* is expected in February 1972. It contains a coded list of some 5,900 commodity class descriptions, combined into sections, divisions and groups and it describes the structure and underlying principles of the classification system.

Standard classification systems form the framework for integrating statistical series derived from different sources. Statistics Canada developed a Standard Commodity Classification during the 1950's with a view to providing comparable commodity statistics from all sources in Canada, but notably for imports, exports, production, shipments and materials used by manufacturers, and transportation (freight). Although the original intent was to use the S.C.C. itself for all commodity surveys, this turned out to be impractical. Instead, working classifications based on the Standard were developed to suit the commodity patterns and collection arrangements found in the different commodity fields. After more than a decade of use, an updated version of the S.C.C. was needed to guide and extend data comparability programs which would be reflected in all commodity statistics for the 1970's.

The revised version is the result of an intensive co-ordinated study, extending over four years, by collectors and users of commodity data within Statistics Canada, supported by frequent consultation with other government departments and the private sector. Apart from the many individual suggestions, the bulk of the recommendations considered in the revision were developed by Statistics Canada research teams composed of subject matter officers, chiefly from the Manufacturing and Primary Industries Division, the External Trade Division and the Input-Output Research Division, under the guidance and direction of Central Classification Staff. Useful comments were also received from other federal government departments, the provinces, trade associations and from representatives of individual companies engaged in manufacturing or trading in commodities.

Volume I – The Classification will be issued first in English. The French version is expected to be available before the end of 1972. *Volume I* is the forerunner to the more detailed companion volumes, *Volume II, Numerical Index,* and *Volume III, Alphabetical Index.* Publication of these latter two volumes, which differ from one another principally in the arrangement of items and which are intended to serve as references, is expected toward the end of 1972. Each manual will contain about 60,000 commodity terms.

Readers interested in more information about the Standard Commodity Classification are invited to contact W. Bokovoy, Central Classification and Company-Establishment Integration, Integration and Development Staffs, Statistics Canada, Ottawa K1A 0V7. Copies of the Revised 1971 Standard Commodity Classification, Volume I – The Classification, catalogue number 12-502, will be available from Publications Distribution, Statistics Canada, Ottawa, K1A 0T6.

Survey of Expenditure at Isolated Posts in Canada

The Prices Division of Statistics Canada has responsibilities for measuring certain limited elements of the living cost differentials encountered by federal government staffs serving at various locations in the more remote parts of Canada, to assist in the establishment and revision of appropriate allowances for such personnel. These isolated posts are situated not only in the Yukon and the Northwest Territories but also in most of the ten provinces.

Early in 1972, in connection with this program, the Comparative Living Cost Section is planning for the first time to conduct an Expenditure Survey by mail to obtain information on spending patterns of federal government employees who served at isolated posts during 1971. This survey is intended to provide a better understanding of conditions at these remote locations, and to enable comparisons to be drawn between spending patterns there and those in other parts of Canada. It parallels a somewhat similar survey undertaken successfully by the International Prices Section two years ago among government employees stationed abroad.

The present project consists of three parts:

(1) Expenditure Survey – A recall survey of expenditures over the entire budget during 1971, broken down under approximately 50 spending categories.

(2) Diaries of Food and Other Household Purchases – Two diaries for completion in two successive 7-day periods early in 1972, to provide details of purchases of food for home consumption as well as a range of household supply items.

(3) A Survey of Supply Sources and Shipping Methods – A questionnaire to yield information about where purchases are made and to indicate how any goods which are not bought from local suppliers are shipped to the isolated post.

The results of this survey, which will provide the first comprehensive information on the spending habits of a sizeable segment of the population in geographically remote communities, are expected to be useful for the improvement of measurements of living cost differentials being encountered by government staffs serving at such locations throughout Canada.

For more information on this survey, contact H. Segal, Assistant Director, Retail Prices and Living Costs, Prices Division, Economic Statistics Branch, Statistics Canada, Ottawa, K1A 0T6.

Canadian International Trade Classification Commodity Index

The Statistics Canada External Trade Division is now in the final stages of producing a new working document for importers and their agents, entitled the *Canadian International Trade Classification Commodity Index.* The need for this index grows out of the decision of the Department of National Revenue, Customs and Excise to apply electronic data processing to the clearance of imported goods. This automation program is being developed in many of its aspects in co-operation with Statistics Canada.

In order to explain the new Commodity Index and its application, it is necessary first to describe the present system of compiling import statistics. External Trade now receives Customs entries and invoices from Customs ports across Canada. These documents, comprising some 10 million a year, are scrutinized, classified and processed by the staff of the Division. The information extracted is assembled to produce import statistics in terms of country of origin, port of clearance, quantity, value, and commodity. The present Import Commodity Classification, made up of some 2,700 individual commodity classes, is used in the monthly report *Imports by Commodities*, catalogue number 65-007, and in other quarterly and annual publications of the External Trade Division.

The users of the present five-digit Import Commodity Classification will find that the new Canadian International Trade Classification Commodity Index is an expansion, containing some 14,000 seven-digit commodity codes for use by importers in completing the redesigned Canada Customs Import Entry form, B-3, as of April 1, 1972. The extension by two additional digits will not only provide greater detail but also facilitate the administrative procedures connected with the clearance of goods through Customs ports across Canada.

The sources of this greater detail were many. Some of the 2,700 classes were sub-divided following conversations with industry, trade associations and government departments in the product areas concerned. External Trade Division also drew on the observations of the divisional commodity specialists of the patterns of trade and kinds of goods traded. For example in some areas, such as chemicals, a large number of trade names and proprietary descriptions were included since the goods are documented in trade according to these rather than by some generic chemical description.

As an example of this exercise in commodity code refinement, class 703-25 "Thermometers", representing about $2 million in imports in 1970, was expanded to give five new seven-digit classifications, involving eleven product descriptions as follows:

703-25 THERMOMETERS	
THERMOMETERS, CLINICAL	703-25-11
THERMOMETERS, CALORIMETRY	703-25-21
THERMOMETERS, LABORATORY	703-25-21
THERMOMETERS, HOUSEHOLD	703-25-31
THERMOMETERS, WINDOW	703-25-31
METERS, TEMPERATURE, ELECTRIC	703-25-41
THERMOMETERS, INFRA-RED RADIATION	703-25-89
THERMOMETERS, RESISTANCE	703-25-89
THERMOMETERS, THERMOGRAPH	703-25-89
OVEN TESTERS, TEMPERATURE	703-25-89
TESTERS, OVEN/TEMPERATURE	703-25-89

Every other commodity class has been expanded in a similar fashion, resulting in some 14,000 new seven-digit commodity classifications applicable to about 45,000 separate commodity descriptions. In the spring of 1972, this detailed commodity classification will be published, in both official languages, as a commodity code index arranged alphabetically to facilitate the preparing and processing of the new Customs entry forms. The present plans call for the revision of the work on an annual basis, making any necessary improvements and including new products coming into Canadian external trade.

Considerable benefits are expected from the new system. Statistics Canada will be able to collect import commodity data in much finer detail, with the resultant improvement in the quality of the commodity data produced. For Customs, the benefits lie principally in savings of administrative costs in the collection of the appropriate duties and taxes. The importers will find that in 1974, once the system is operating nation-wide, they will be able to obtain release of their goods more quickly. As users of Statistics Canada statistics, they will also benefit from expanded and more timely data on the products in which they are interested.

For further information, please contact A.J. Wibe, Chief, Commodity Intelligence Section, External Trade Division, Economic Statistics Branch, Statistics Canada, Ottawa, K1A 0V5.

Work Continuing on Machinery and Equipment Price Indexes.

In June 1971, the Prices Division of Statistics Canada first published price indexes of construction machinery and equipment *(Prices and Price Indexes,* catalogue number 62-002). This marks the first of a series of industry-classified price indexes relating to purchases of machinery and equipment. By 1972, it is expected that similar indexes for machinery and equipment will be available for the following industries: forestry, truck transport, sawmills, road maintenance, open pit mining, and storage and warehousing.

These indexes represent an expansion in the availability of detailed commodity indexes for complex goods. Such indexes are of use to a large number of government and private analysts for the price correction of historical value series or for forecasting economic trends. Manufacturers of such machinery and equipment, purchasing agents and market researchers can also make similar use of these indexes.

To increase their usefulness, price indexes for Canadian-made machinery, imported machinery and their composite indexes will be published wherever possible for the commodities in the sample and for the total industry, on a monthly basis in *Prices and Price Indexes.* The manufacturers' selling prices used in these series are be adjusted as necessary for exchange rates, duty rates and federal sales taxes.

For more information on this subject, contact J. Plumpton, Industrial Prices Section, Prices Division, Economic Statistics Branch, Statistics Canada, Ottawa, K1A 0T6.

Family Expenditure Surveys: 1969 Results and 1972 Plans

Results: Survey of Family Expenditures, 1969 – It is expected that first results will soon be available from the large-scale survey of family expenditures carried out early in 1970 by the Family Expenditure Section of the Prices Division, Statistics Canada. In this survey, a sample of 22,000 households was assigned for interview to obtain a complete accounting of family expenditures and income in the 1969 calendar year. The sample was designed to represent families and individuals living in private dwellings, urban, rural non-farm and farm, in the ten provinces. In contrast, previous expenditure surveys carried out during the past two decades have been limited to selected urban centres. The 1948-49 survey, although similar in scope to the 1969 survey, yielded usable results only for the non-farm population. This is the first expenditure survey, therefore, to provide an all Canada composite expenditure pattern.

On the basis of experience in earlier surveys, the sample of 22,000 households was assigned with the expectation that it would net approximately 15,000 usable records. In fact, about 15,500 schedules were considered usable. The response rate of 69 per cent was slightly higher than that obtained in the urban surveys of 1964 and 1967, and appreciably better than the 65 per cent response obtained in the national food diary survey conducted during 1969.

A generalized tabulation system has been developed with potential to extract the maximum amount of information from these records. In addition to the complete detail on expenditure, income and other financial changes entered on the 32-page schedule, each family tape record contains forty codes which classify it according to location, urbanization, income, family size and composition, age of head, etc., and other characteristics developed from the expenditure record. In all, each tape record consists of 1,234 information fields.

It is planned to publish this information in four volumes under the general title, *Family Expenditure in Canada, 1969.* The volumes will be entitled as follows:

Volume 1 – All Canada, Urban and Rural, catalogue number 62-535
Volume 2 – Regions, catalogue number 62-536
Volume 3 – Major Urban Centres, catalogue number 62-537
Volume 4 – The Analysis of Family Expenditure, catalogue number 62-538

Standard tables will be similar in format to those published in recent reports for the 1964 and 1967 surveys *(See Statistical Observer, Volume 4, Number 3, October 1971, p. 12).* Expenditures for a given classification will be presented as a summary table showing family characteristics, main expenditure groups and their distribution in the total, and also as a lengthy, detailed table showing average expenditure per family and percentage reporting for individual items. Volumes 1, 2, and 3 will consist mainly of these two standard types of tables, with introductory text. Volume 4 will contain a variety of analytical material including estimates of variance and the results of multiple regression analyses of expenditure data from the 1969, 1967 and 1964 surveys.

The publications will present a selection from a large body of unpublished material. Because the size of the survey makes possible a greater degree and variety of cross-classification than ever before, many tables will appear in summary form only, with additional detail available upon request. The problems of storage and easy access to unpublished tabulations are being met by transferring the major part of computer output directly from tape to microfilm. A microfilm Reader-Printer will facilitate scanning and retrieval to service specific requests.

Although the expected release date of the information resulting from this major survey represents an improvement in timeliness over earlier ones, potential users may wonder why information collected in the first quarter of 1970 cannot be processed and released before the end of the following year. Family expenditure surveys, especially large-scale ones, are time-consuming ventures at all stages of collection, clerical processing, key-punching, computer editing, programming, and even, in relative terms, in the time required for computer processing. No way has yet been found to by-pass the operations of clerical editing and key-punching which are particularly time-consuming for expenditure surveys.

In processing the 1969 survey, most of the clerical editing was completed by the end of 1970, but the last stages, involving

the acceptance or rejection of problem schedules, continued in 1971, with key-punching finally completed in the middle of that year. After repeated runs of large blocks of data through computer edits, the basic data tape was declared "clean" by early September 1971. This allowed the first stages of production, comprising creation of 25 sort tapes, as a preliminary to running tabulations, to begin. A set of tables was run to test the functioning of the program with the complete data set, and to obtain a realistic estimate of tabulation time and cost.

A "one-shot" survey, even when it builds on past experience, cannot achieve the timeliness of repetitive surveys such as the 1969 food expenditure survey in which editing and other processing were carried out cumulatively as the series of monthly surveys progressed. Results for the food expenditure survey, which also referred to the calendar year 1969, were released by the end of 1970. The first volume of the report, *Family Food Expenditure in Canada, 1969*, Statistics Canada catalogue number 62-531, is already available and will be followed by a second volume, catalogue number 62-532.

Timeliness, although important, is not the only consideration in expenditure survey methodology. Record-keeping or "diary" was chosen as the most appropriate method for collecting detail on frequently purchased items such as food and other household supplies. It has also been used in the United Kingdom and other European countries for the collection of all household expenses, supplemented by recall information for expenses made on an annual basis. The method of recall interview, as used in Canada and in the large decennial surveys conducted in the past by the United States Bureau of Labor Statistics, involves a lengthy, comprehensive interview, with references wherever possible to records, checkbook stubs, etc. It is subject to criticism because of the proven fallibility of human memory. Researchers in survey methods who have undertaken to measure memory loss have shown that it generally increases with the length of the recall period. On the other hand, record-keeping, in which memory loss is minimized, requires continued co-operation and carries the risk of altering or conditioning the behaviour of respondents by the mere fact of survey participation. Higher first-week expenditure is a generally-observed characteristic in record-keeping surveys of two weeks or more. For food and relatively short-term consumption goods, it may be assumed that any unusual spending in the first week is compensated in the second, and that the two weeks together form an acceptable estimate for a two-week period. For other items, this assumption may be less valid.

Plans for Surveys in 1972-73 – The intention of the foregoing remarks is not to deal exhaustively with methodology problems, but to indicate that such problems exist. Because of the urgency of data needs, the systematic exploration of data collection problems frequently receives low priority. However, in 1972, with the extremely large body of 1969 data becoming available, it has been decided to incorporate variations in method into the survey program, in conjunction with an up-dating of certain areas of the budget. Survey activities in 1972 are being concentrated on two partial budget surveys. The first of these surveys, which will be carried out in February and March of 1972, with reference to the year 1971, covers expenditure on shelter, home furnishings and other household durables, including vehicles. Purchase data will be collected in the same manner as in 1969, using an inventory approach to household appliances and vehicles. Apart from the partial coverage of the budget, the experimental feature of the survey consists of detailed questions on credit purchases, designed to elicit more complete information on financing costs and interest charges than has been collected in past surveys. A sample of 4,600 households, expected to yield more than 3,000 records, is being assigned in eight major urban centres.

The second survey will be carried out as a bi-monthly series from March 1972 to January 1973, and will attempt to collect full detail on clothing purchases by recall for the two preceding months. In each bi-monthly survey, a sample of approximately 1,650 households will be interviewed, totaling almost 10,000 households over the six surveys. Both the foregoing surveys will be carried out in the urban centres of St. John's, Halifax, Montreal, Ottawa, Toronto, Winnipeg, Edmonton and Vancouver.

Commencing in 1972, more resources will be committed to research in survey methods. It is hoped also to draw some useful conclusions from a new approach in survey methods being undertaken in the United States by the Bureau of Census, which now has the responsibility of carrying out expenditure surveys for the Bureau of Labor Statistics. This survey, which will replace the traditional decennial survey carried out by the BLS, will be a combination of quarterly recall and a diary survey, each administered to a different sample. For the quarterly survey, a panel of families from an assigned sample of 20,000 to 25,000 households will be interviewed five times, beginning in the first quarter of 1972 and ending the first quarter of 1973. This survey will not cover the total budget, but will be supplemented by the diary survey which will cover all household purchases for a two-week period, with an expected sample of about 17,000. Although there has been some testing of both panel and diary methods, misgivings have been expressed by data users as well as expenditure survey experts concerning the response results in the panel tests, the complete break in continuity with past surveys, and the formidable task of integrating the two sets of data. A session of the American Statistical Association annual meeting in August, 1971 was enlivened by a discussion under the topic *Methodology of Consumer Expenditure Surveys*, which indicated again the problems and challenges of consumer expenditure surveys. *Inquiries about Family Expenditure Surveys may be directed to I. McWhinney, Chief, Family Expenditure Section, Prices Division, Economic Statistics Branch, Statistics Canada, Ottawa, K1A 0T6.*

Export Promotion Seminar

In June 1971, Statistics Canada played host to 24 business and government officials from seventeen emerging nations for a one-day seminar organized by the External Trade Division. This was the second time the bureau has participated in the Export Promotion Program, organized by Waterloo Lutheran University and sponsored by the Canadian International Development Agency. The visitors represented countries in Africa, South East Asia, and the Caribbean.

On hand to welcome the group was V.R. Berlinguette, Director General, Economic Statistics Branch, who also introduced the day's speakers – John Wall, Chief, Trade Information Section, External Trade, Andrew Billingsley, Statistics Use Development Officer, Ottawa Region, and Art Wibe, Chief, Commodity Intelligence Section, External Trade.

Mr. Wall described Canada's trade statistics, and how to use them in assessing market opportunities in this country. The theme of marketing research was continued by Mr. Billingsley who described how other Statistics Canada reports could be used to complete the market assessment. Mr. Wibe completed the formal presentations, outlining the classification system used in the bureau.

The visitors had been divided into four study groups by the University and a marketing study assigned to each. In the time remaining, everyone took advantage of the opportunity of doing some research on these studies. Several commodity officers from the External Trade Division sat in on the discussions, providing assistance where necessary.

The second half of the delegates' Ottawa visit was spent with Department of Industry, Trade and Commerce officials.

Grain Statistics Symposium

The Agriculture Division and the Statistics Use Development staff of Statistics Canada, in co-operation with the Canadian Wheat Board and the Canadian Grain Commission, organized a symposium on grain statistics to increase communication between the various suppliers of grain and crop statistics and between the suppliers and users of these data. The Agriculture Division regarded this symposium as the first phase in developing a co-ordinated approach to the application of newer techniques to better serve the agriculture sector, and the identification of problem areas in grain statistics.

The two-day meeting in Winnipeg was attended by representatives of federal and provincial governments, agri-business and farm organizations. L.E. Rowebottom, Assistant Chief Statistician, Statistics Canada and M.H. Head, General Director, Management Information Services Division, Canadian Wheat Board, opened the conference. At the meeting, papers were presented describing the various roles of the Agriculture Division of Statistics Canada, the provincial agricultural statistical agencies and the Canadian Grain Commission in providing grain statistics. The statistical requirements of the grain trade, of agriculture extension specialists, and of those concerned with policy formulation were also outlined.

During the meeting, there was opportunity for general discussion of the papers. As well, the representatives divided into four discussion groups to deal in more detail with some specific aspects of grain statistics.

A number of topics arising from these discussions were summarized and presented to the delegates at the close of the symposium. The first recommendation was to set up a standing committee on grain statistics which would work toward identifying and solving problems in crop and grain statistics. This committee would be composed of representatives from statistics-producing agencies and user groups.

In addition, the symposium suggested that the various agencies involved in producing grain statistics (Statistics Canada, the Canadian Grain Commission, the private grain trade, etc.) should attempt to improve co-ordination of their activities in order to provide better service to statistics users.

Many specific areas of improvement in grain statistics were suggested by the delegates, such as: more small-area data; more climatological data; information on the movement and pricing of non-quota grain, especially feed grains; the publication of an expanded package of data on farm management practices; information on the supply and pricing of agricultural inputs; more detailed statistics on Durum wheat; clarification of the basis on which dockage is calculated in grain production estimates; a break-out of the category "animal feed, waste and dockage" in supply-disposition tables for grains; and the use of metric units in publishing statistics for grains traded internationally.

The delegates also had a number of suggestions concerning the publication of statistics, especially with regard to improved timeliness. Interest was shown in the use of CANSIM in this area. The need was expressed for a better market intelligence system to regularly provide primary producers with timely, high quality market data. The expanded use of the news media as vehicles for releasing data quickly was also suggested.

The Agriculture Division of Statistics Canada has already acted on some recommendations of the symposium and is considering others. For example, metric units were used in the November 1971 production estimates, in addition to the traditional units of measurement. This practice will be continued in the next release of export data. Also, preparations are underway for setting up a standing committee on grain statistics, and the first meeting is expected to be held in the spring of 1972. *Readers interested in more information on this symposium are invited to contact R. Johnson, Crops Section, Agriculture Division, Socio-Economic Statistics Branch, Statistics Canada, Ottawa, K1A 0L7.*

NEW REPORTS

New Publication from Labour Division

To meet the growing demand for more information in the specialized field of labour statistics, the Labour Division of Statistics Canada's Economic Statistics Branch has produced a new publication, *Notes on Labour Statistics.* This publication is designed to make available the results of analytical and developmental work in labour statistics, especially that information relating to the various labour data series produced in Statistics Canada.

The 50- to 75-page journal will include articles and notes on such topics as the labour force, employment, mobility, earnings, hours of work, job vacancies, pensions, etc. The material will be supplied mainly by the staff of the Labour Division but contributions will also be considered from labour economists in other government departments and agencies (both federal and provincial) and from those in the academic community.

Initially, *Notes on Labour Statistics* will be an annual publication, although it may be issued more frequently in the future. Main users of the information in this journal are expected to be government and university economists and statisticians, as well as executives in business and labour organizations.

The first issue of *Notes on Labour Statistics* is planned for release in February 1972. Along with technical articles, this issue will contain a review of recent developments in labour statistics in the bureau, describing new work in progress in the Labour Force Survey, the Job Vacancy Survey and in other areas such as pensions, labour costs and employment by occupation.

Topics of the technical articles appearing in the first issue are: the relationship of earnings of males and females in the manufacturing sector; national industrial accident statistics; a discussion on the use of Job Vacancy Survey data in labour market analysis; employment and unemployment of youth in summer months; seasonal patterns in part-time work; job search patterns; youth participation in the labour force; and the educational attainment of the Canadian labour force.

Copies of Notes on Labour Statistics, catalogue number 72-207, are available from the Publications Distribution Unit, Statistics Canada, Ottawa, K1A 0T6. Inquiries about material in the publication may be directed to H. Buckley, Co-ordinator, Manpower Research and Development, Labour Division, Economic Statistics Branch, Ottawa, K1A 0V1.

1971 Edition of U.S. Statistical Abstract

The Statistical Abstract of the United States is the standard summary of statistics on the social, political and economic organization of the United States. The 1971 Abstract, the 93rd edition of this valuable reference book, was recently released by the U.S. Bureau of the Census. Special efforts were made to incorporate 1970 Census information into this edition.

The 1971 Abstract includes 1970 Census figures for total population; population density; population by metropolitan, urban and rural residence; and such characteristics as sex, age and race at national and state levels. Basic 1970 population data also are presented for Standard Metropolitan Statistical Areas and for large cities.

Data of current significance are contained in the new tables on minority groups, police officers killed, jails and jail inmates, credit card banking, pesticides, characteristics of college faculty members, waste paper utilization and price indexes for selected countries.

The Statistical Abstract of the United States: 1971 is available from the Superintendent of Documents, United States Government Printing Office, Washington, D.C. 20402, for $5.50.

Nova Scotia Statistical Review

The Voluntary Economic Planning Division of the Nova Scotia Department of Development has released its third annual statistical review. This publication, prepared by the economic analysis section of the division, is designed for use by people in government and industry who require concise and readily available information on the performance of the Nova Scotia economy, and its position compared with the selected provinces and Canada as a whole.

The majority of the statistics have been extracted from various Statistics Canada publications. All sectors of the economy have been covered – primary industries, manufacturing, services, trade, etc. In addition, there are statistical series on population, income, the labour force and capital investment.

Copies of the Nova Scotia Statistical Review are available from the Department of Development, Halifax, Nova Scotia.

Canadian Aviation Handbook

In recent months, the Aviation Statistics Centre of the Transportation and Public Utilities Division has been putting the finishing touches on a new publication, *Aviation in Canada, 1971.* Compiled and written by personnel of the Centre, the publication was designed to provide a broad, interesting view of the growth and development of the air transport industry and some of the problems which have occurred. In addition, it is valuable reference source on all aspects of flying in our country.

For the interested reader, four chapters deal particularly with historical facts, the roles of international organizations and the federal and provincial governments in aviation, and numerous other items of information such as the provincial distribution of flying, gliding, and aircraft safety, since the turn of the century. The Aircraft Technology chapter points out Canadian research, development and manufacturing of aircraft, imports and exports and the operating characteristics of different planes in this country, concluding with a review of STOL technology. Three sections outline specifically the growth of airport traffic, the flow of passengers and cargo on the main routes and the operations of commercial air carriers, emphasizing the activities during the decade of the 1960's.

The handbook shows many statistical tables originally compiled for internal studies which now will be released to the

general public. It contains data on government expenditures for aviation and explains how air traffic control works. Tables and charts illustrate the increase in registered aircraft since World War II and the rise in the number of licensed pilots, numbering 40,000 in 1971. Each chapter is documented with tables and charts, and a glossary of terms is appended to the book for the better understanding of specific aeronautical terms.

The value of *Aviation in Canada, 1971* is that it contains, in one volume, the statistical highlights from the last ten years, drawn from Aviation Statistics Centre publications, as well as providing the general background information required for total view of the air transport industry in Canada.

For additional information, contact Jan Bekooy, the Aviation Statistics Centre, Transportation and Public Utilities Division, Economic Statistics Branch, Statistics Canada, Congill Building, 275 Slater Street, Ottawa, K1A 0N9.

Canada's International Investment Position

This major new report, published in December, brings together available data on Canada's international investment position, extending and revising material published in *Canada's International Investment Position, 1926 to 1954*, catalogue number 67-503, and subsequent releases in Canada's Balance of International Payments reports. The main summary tables contain data from 1926 to 1969, although most of the detail extends only to 1967. In addition to statistics on the year-end levels of Canada's international assets and liabilities, the report contains tables covering investment income flows and broad measures of the extent of foreign ownership and control in the Canadian economy. The principal statistical changes in this latest report are the increased geographic detail of foreign direct investment in Canada and both the geographic and industrial detail of Canadian direct investment abroad. The latter series provide, for the first time, data covering numbers of investors and investments, and size distributions. With a special focus on Canada's direct investment in developing countries, more detailed information than hitherto available is also given for Canadian export credits.

Copies of Canada's International Investment Position, 1926-1967, catalogue number 67-202, are available from the Publications Distribution Unit, Statistics Canada, Ottawa, K1A 0T6.

Inquiries about material in the publication should be directed to D.K. McAlister, Chief, Balance of Payments Section, Balance of Payments and Financial Flows Division, Economic Accounts Branch, Statistics Canada, Ottawa, K1A 0Z8.

Director Retires

F.F. Harris, Director of the Health and Welfare Division of Statistics Canada's Socio-Economic Statistics Branch has retired after 22 years with the bureau. In 1949, Mr. Harris came to Statistics Canada from the Newfoundland government service.

As well as developing many new programs in the Health and Welfare Division during his career, Mr. Harris is well known internationally. He was involved in both the seventh and eighth revision of the World Health Organization's international classification of diseases. This classification is used throughout the world as a guide in producing health statistics.

Appointments

Norman F. Beaudoin has been appointed Head of the Aggregate Productivity Measures Unit of the Statistics Canada National Output and Productivity Division.

C.H. Bubeck, formerly Capital Markets Sector Head, Financial Flows Section of the Balance of Payments and Financial Flows Division, was promoted to Chief of the Provincial Government Section, Governments Division, Statistics Canada.

D.W. Hall has been named Director of Personnel Administration for Statistics Canada. Mr. Hall has worked in a wide variety of personnel positions, including the federal departments of Public Works and Agriculture, the Public Service Commission and the Canadian International Development Agency. His last position was with the Personnel Policy Branch of the Treasury Board Secretariat.

G. Phelan has been appointed Head Librarian of the Statistics Canada Library. Miss Phelan's experience in library work includes positions with the Bank of Montreal Library and various posts in the McGill University Libraries system.