Specific Problems

The current system has been operational in much the same general form for about ten years. During this time a number of problems have been identified, some of which are listed below.

Coordination between the BRMF and the various survey operations is not tightly regulated. The quality of some frame data items is poor; multi-establishment structures are not current.

Coverage via the BRMF is not complete as non-employers are excluded, and there is no linkage to income tax files which could provide such data. This presents problems for surveys where self-employed individuals contribute substantially to industry totals, e.g., retail trade, services, construction trades.

The data requirements of a comprehensive small business program for which there is increasing need cannot be fully met.

Recent Developments

Some of these problems have been addressed in developments over the past few years as the following examples illustrate.

To improve quality, control procedures have been introduced for the major clerical operation. The first phase of a data quality assurance scheme for BRMF data has been implemented.

To improve data sharing, protocols for exchange of data between the BRMF and individual survey data files have been developed.

To obtain coverage of non-employers and, in addition, to facilitate the use of tax data in lieu of survey data, the annual "Census" of Construction (actually a sample survey) employs multiple frame techniques based on income tax data.

To cope with inactive units on the BRMF, the new monthly Survey of Employment, Payrolls and Hours incorporates a novel procedure which permits such units to be deleted from the sample while retaining the information from them for estimation purposes.

Future Developments

Some substantive new developments are in the pipeline.

Afar reaching new proposal suggests that the current employer data bases for the BRMF be replaced by income tax data. This would provide a list of employers and non-employers therefore meeting the needs of all annual surveys, along the same lines as the present Census of Construction, and at the same time better supporting a small area business program. Under this arrangement sub-annual surveys would be linked to annual benchmarks based on tax data.

Enterprise identifiers from the Intercorporate Ownership program will be included on the BRMF in the near future. This will greatly facilitate establishment profiling, which may be further improved by the introduction of a self-profiling program for the smaller multi-establishments.

The 1980 SIC is being implemented. This will standardize industrial classification and procedures for delineation of establishments.

Generalized software is being developed for
The BRMF to coordinate all company contacts and to provide a general vehicle for collection of (frame) data.

These points are all further elaborated in later sections of the paper.

**CONCEPTS**

**Economic Activity and Its Measurement**

The scope of economic surveys is market-based activity, i.e. "activity associated with the production of goods and services and the sale of those goods and services in final markets" [1, p. 29]. In this context, "market-based" is defined "to refer to the money exchange economy which includes the activities of governments and non-profit making organizations as well as the activities of business enterprises operating for a profit" [1, p. 28].

The System of National Accounts (SNA) strongly influences the objectives of individual economic surveys by providing a common theme and a more or less unified set of concepts, definitions and classification systems. The central component of the SNA is the National Income and Expenditure Accounts which generate "a basic statistical picture of the key economic processes of production, income generation, sales to major markets, consumption and capital formation" [1, p. 29].

Perhaps the single most significant measurement notion is that of economic production. All the various forms of economic activity may be encompassed within a unified conceptual framework which is keyed to and built around the central concept of economic production [1, p. 62]. The major "production account" variables may be summarized as: gross outputs, expenditures, inventories, employment, salaries, wages and supplementary income, depreciation and net operating surplus [2, p. 16].

**Industrial Classification, Statistical and Reporting Units**

"It is desirable to be able to isolate meaningful types of specialization among producers and this is done by breaking down the total of economic production into industries; that is, groups of producing units engaged in similar types of activity in relation to goods and services" [2, p. 11].

There are many possible alternative classification systems depending upon the unit upon which the classification is to be based. The single most important scheme is the Standard Industrial Classification (SIC) of establishments. The "establishment" (as a statistical unit) is defined [2, p. 16] as "the smallest operating entity which produces as homogeneous a set of goods and services as possible and for which records provide data on the value of output together with the cost of materials used and the cost and quality of labour resources employed to produce the output and for which records or estimated allocation can provide the full range of production account variables to calculate value added."

It follows from the definition that the data items specifically associated with an establishment are SIC and production account variables. In addition, it is useful to associate a means of identification (name, address), a location and measures of size.

The "establishment as a reporting unit," or, more briefly, "reporting unit," is the unit from which the "principal" production account statistics can actually be obtained. It is a practical compromise of the target statistical unit.

Delineation, i.e., identification, of establishments and reporting units is referred to as "profiling." Establishments may be "single" or part of multi-establishment enterprises.

Division of establishments by industry into mutually exclusive groups is the basis for specification of the important individual "subject matter" annual surveys of principal statistics.

Ideally, all economic surveys including subannual surveys and surveys for other data should be based on the same set of units. In practice, a demand for finer geographic and/or industrial disaggregation is sometimes recognized. Also for certain surveys, e.g., aviation, there is a regulatory process which determines the most practical unit.

**Survey Frame Data**

The frame data requirements for any individual economic survey may be summarized as comprising a list of the appropriate statistical and reporting units together with the following data items:

(a) identification number;
(b) contact and other operational information - name, address, contact method, follow-up procedure;
(c) classification information - SIC, size measure(s), geographic code;
(d) linkage information - to other units in the list, to past lists, to related lists; and
(e) maintenance information - date of birth of unit, of record, source and date of last update, etc.

**Use of Administrative Data**

Lists of establishments and other statistical and reporting units have, in the first instance, to be built. Generally the most cost effective construction method is to make use of existing lists created for other, usually administrative, purposes.

Once a survey frame has been constructed the most important source of frame maintenance information is likely to be the survey operation itself. However as the data thus collected refer only to units currently in the survey sample they must be supplemented from other sources. Even a census operation requires updating information concerning potential "births." Again administrative sources are utilized.

Use of any one administrative list will introduce certain problems due to the fact that the list is designed for the administrative, not the statistical, function. There may be problems in linkage. There may be slight differences in concepts, timing problems, missing data items, unsatisfactory quality, etc. Use of two or more unrelated administrative sources will tend to compound these problems; linkage and unduplication are likely to be particularly difficult. The utility of each must also be judged in terms of cost of acquisition and processing.

In constructing lists of statistical units for economic surveys from diverse sources, two unifying concepts which can facilitate coverage and unduplication are those of "legal entity" and of
"enterprise" (defined on the basis of 50% ownership or control). All economic production can in principle be related to an enterprise which is itself a legal entity or group of legal entities. Thus the set of legal entities provides the basis for complete coverage.

Central List and System

A central repository for frame data is highly desirable for several related reasons:
(a) to enable control of coverage, e.g., the division of the set of establishments into subject matter survey subsets which are mutually exclusive, and the identification of industries not covered by any survey;
(b) to enable efficient transfer of data between surveys and, especially, to be a focal point for receipt and storage of administrative data;
(c) to provide standardization and coordination of frame data acquisition and usage; and
(d) to carry out functions, e.g., multi-establishment profiling, control of response burden, which cross survey boundaries.

The disadvantages stem mainly from the size of such a system, its inertia and the corresponding lack of access and control which individual survey managers may then feel. The practical as opposed to theoretical value of a central function depends upon the service actually provided and the willingness, voluntary or mandatory, of survey managers to make use of that service.

CURRENT PROVISION OF FRAME DATA

Introduction

A partial list of economic surveys is appended which serves to demonstrate the broad scope of the program and the breakdown by subject matter Division. The Divisions and the survey sections within each Division have a considerable degree of autonomy. In particular, survey managers are largely responsible for their own frames and tend to utilize centrally provided frame information to the extent their judgement and resources allow. Thus the acquisition of frame data varies considerably from one survey and Division to another, as will be briefly illustrated in the following paragraphs.

The current system by means of which frame information is centrally provided to the various economic surveys in fact comprises four virtually independent subsystems. The principal component is the Business Register Master File (BRMF) and the other components are the income tax data and the intercorporate ownership data files.

BRMF: Objectives

The BRMF was implemented in 1972. It was accompanied by a policy statement [3] defining the five principal objectives of the operation as:
(a) control of economic survey coverage;
(b) control of classification;
(c) implementation of a common identifier;
(d) maintenance of an inventory of surveys for which each unit is in scope / in sample; and
(e) tracking changes of units over time.

Thus the original intention was that the BRMF should be a complete, up-to-date list containing frame data for all economic surveys. In practice it has not fully achieved this role.

BRMF Implementation Based on Employer Data

The BRMF is currently a set of about 1.3 million legal entities, establishments and employer account records. It is based primarily on employer data from Revenue Canada (RCT).

The essential features of employer data in this context are as follows.
(a) The data refer exclusively and exhaustively to "employers" required by RCT to permit payroll deductions.
(b) The data are provided free of charge by RCT.
(c) Data available monthly in machine-readable form provide updates to employer account name, address and remittance status, hence can indicate additional or changed economic production but are not by themselves adequate for delineation or classification of statistical or reporting units.

(d) Data available annually in machine-readable form include total earnings and number of employers paid via each account during the year.

(e) Copies of (PD20) questionnaires sent by RCT to new account holders are forwarded to the BRMF. A new account cannot be taken as an indication of the birth of a unit because existing units can open new accounts. Data for perhaps 70% of all new accounts are received within six months of the account being opened. The data are generally sufficient for the identification and approximate classification of small statistical unit births, but must be supplemented by research when classification information is inadequate or when the birth of a large unit (20 or more employees) is indicated.

Employer data generate about 180,000 potential births and 1,000,000 updates per annum. This information is available to all surveys but in general only the birth data are used.

Major surveys requiring specific use of BRMF data for frame maintenance include the Census of Manufactures, the Motor Carrier Freight Survey, the Retail Trade Survey and the Survey of Employment, Payrolls and Hours. Each of these surveys maintains its own frame file incorporating information received from the BRMF and from the survey operation itself over the years. The Retail Trade Survey Frame is supplemented by an area file to provide coverage of non-employers.

The BRMF is the repository for profiling information about multiple establishments and is the principal instrument for ensuring proper subdivision of the universe by subject matter.

Income Tax Data

The second and third sources of frame information are the corporate (Form T2) and individual (Form T1) income tax data from Revenue Canada (RCT). These data are received by two operations within Statistics Canada. They are not systematically linked to the BRMF.

The essential features of corporate tax data from the viewpoint of frame maintenance are as follows.
(a) The data refer exclusively and exhaustively to all Canadian corporations and licensed branches of foreign corporations required by RCT to report income.
(b) Data available in machine-readable form from RCT for all tax filers include corporate name, address and major financial statistics (assets, sales, taxable income, profits and equity); these data are not sufficient for delineation or SIC coding of statistical
units.

(c) RCT provide access to copies of T2 forms and accompanying schedules and financial statements from which more detailed data may be transcribed; delineation and SIC coding of establishments is possible for corporations with non-complex operations; sometimes the data are inadequate for assignment of a detailed SIC code.

(d) The data are available from 8 - 18 months after the reference calendar year and refer to a fiscal year ending within the reference year.

All active corporations (about 450,000) are assigned an SIC code by Statistics Canada. Various other data items are transcribed for samples of tax returns in two separate operations. A sample of about 25,000 mostly large corporations is used for the Annual Survey of Corporations Financial Statements. Data for a second sample of about 40,000 smaller corporations are made available to subject matter surveys.

The essential features of individual (Form T1) tax data are as follows.

(a) The data refer exclusively and exhaustively to individuals required by RCT to file an annual income tax return.

(b) Data available in machine-readable form from RCT for all tax filers include (RCT) name and address, and gross and net self-employed revenues in five broad categories; these data are not sufficient for delineation nor SIC coding of statistical units.

(c) RCT provide access to the original T1 forms and accompanying documents; a sample is copied, SIC coded and various data items are transcribed; these data enable delineation and classification of statistical units except that partnership relationships are not explicitly available and comprehensive classification is not always possible.

(d) The data are, in essence, available within 1 year of the reference year; the financial items refer to fiscal years ending within the reference year as chosen by the tax filers.

Data for a sample of about 160,000 individual tax returns are transcribed each year.

In summary, corporate and individual tax data are made available to all economic survey operations. They are put to use in a variety of more or less comprehensive ways. In the Census of Construction tax data supplement the large business frame and are used in lieu of some survey data for small businesses. For the Census of Manufactures they are used in lieu of survey data but not for frame maintenance; for surface transportation surveys the reverse is the case. A full program for comprehensive tax data usage is being developed for Merchandising and Services surveys.

CALURA Data

The other major source of frame information is the intercorporate ownership data collected under the Corporations and Labour Unions Returns Act (CALURA). These data provide comprehensive ownership information presently covering some 50,000 corporations.

The data are not linked to the BRMF and are not systematically exploited by other economic surveys for frame maintenance purposes.

Other Data Sources

A variety of other administrative data sources are available for frame maintenance purposes. Some are industry specific, for example arising from a regulatory process, and are tailor-made for survey of that industry. Aviation statistics are an example. Other sources such as telephone directories, new corporate charters and charter amendments are used on an ad hoc basis.

PROBLEMS

The current system for central provision of frame information has been operational in the same general form for about two years. A number of problems which have been identified, several of which stem from the fragmented nature of the system, are outlined in the following paragraphs.

Policy, Concepts, Standards

Despite the BRMF policy document [3] there are differing perceptions of the role a central function should play in the provision of frame data to surveys.

Each individual survey operation maintains its own frame, generally using the BRMF for birth information but otherwise, to a lesser or greater extent, independently. This results in considerable duplication and possible inconsistency of data, and some duplication of effort.

Frame data ("nature of business") enquiries are not fully coordinated across all surveys, and survey feedback to the BRMF is not comprehensive.

There is a perception that some of the resources spent on maintenance of BRMF smaller units could be better devoted to improving the quality of larger units, especially multi-establishments.

Both 1960 and 1970 versions of the SIC are in current use and there are unofficial disaggregations and combinations. Thus, in practice, industrial classification is not yet fully standardized. There are also variations in interpretation of the establishment definition in terms of geographic breakdown and the treatment of ancillary activities.

Multi-Establishment Structures

The resources devoted to the maintenance of multi-establishment structures have been minimal for several years. Also these structures are not fully defined at the enterprise level as intercorporate ownership data are not presently held nor linked to the BRMF.

Coverage

The BRMF is not complete as non-employers are virtually excluded and there is no linkage to income tax files which can provide such coverage. This is a problem for surveys which, on the one hand, require current frame data such as are provided by the BRMF, whilst, on the other hand, require coverage of non-employer businesses. (The latter make a significant contribution in certain industries, e.g., services, construction trades, retail trade). In particular the BRMF cannot fully support a small area business program for which there appears to be ever-increasing pressure from statistical data users.

Data Item Quality

Although there is quality control for certain clerical operations performed in the collection and capture of frame data there has been no comprehensive quality assurance program. Ad hoc assessments of quality, usually in situations where some particular problem has been encoun-
tered, have lead survey managers to have considerable reservations about the reliability of data provided from central files. Specific examples cited are that industrial classification is out of date or not available at the required level of detail, that required size measures are imprecise or not available on the BRMF, and that BRMF units flagged as being in active economic production are in fact inactive or duplicates.

Computer Systems

The BRMF data base system which was put into production over the period 1979-81 incorporates many of the features, including identification numbering, of the original tape based system. It is not sufficiently flexible to incorporate substantial enhancements, for example, the full introduction of enterprise data.

The system for maintenance of corporate tax data is designed primarily to service the Annual Survey of Corporations not the full range of surveys. The system for individual tax data provides a historical base; it was not designed for two-way data exchange with surveys.

RECENT DEVELOPMENTS

Some of the problems described in Section 4 have been more or less comprehensively addressed by methodological, operational and systems developments over the past few years.

Policy, Concepts and Standards

A system was developed some years ago for the Census of Construction (COC) to provide full coverage of non-employers and to facilitate use of income tax data in lieu of survey data [7, 8]. The COC is, in fact, a sample survey despite its name. It is based on the universe of corporate and individual tax filers, not employer data, and is considered to be, in some sense, a model for combined use of income tax and employer data. Larger establishments (revenues of $500,000 or more) are identified and covered by a full scale census questionnaire. Data for smaller units are obtained by sampling tax returns, supplemented by a sub-sample of non-financial items collected on a short survey questionnaire. The corresponding multiple frame techniques have recently been revised to incorporate a new procedure [9]. BRMF data are used to identify new, potentially large, establishments and to control coverage of the census portion of the universe, in particular, the partitioning of multi-establishment corporations to the various survey operations of which COC is one.

Data Item Quality

Formal quality control (QC) procedures using acceptance sampling have been introduced for the basic BRMF clerical operation (processing PD20 forms for new employer accounts).

For several years a comprehensive QC scheme has been in place for data transcription and SIC coding of the smaller tax returns. More recently QC has been introduced for industrial classification of all corporations.

The first phase of a scheme [10] to measure and control BRMF data quality for single establishments has been introduced. Preliminary results indicate that about 20% of all BRMF units flagged as active are in fact "dead" (inactive or non-existent); about 8% are misclassified by industry division and 20% are misclassified at full (3 digit) SIC level; about 7% of the contact names and addresses are inappropriate, etc. [11]. Measures for the introduction for the Survey of Employment, Payrolls and Hours are obtained from the BRMF. Each month new units are rotated into sample as part of survey operations. Specific procedures have been built into the design to cope with the high proportion of dead units encountered in the sample. Retention of such units in the sample throughout the rotation period would be inefficient but their straightforward replacement by new units would cause subsequent bias. The procedure adopted [12] computes and maintains an estimate of the number of dead units in the universe whilst permitting such units to be dropped from the sample.

CURRENT AND FUTURE DEVELOPMENTS

Policy, Concepts, Standards

A far reaching proposal has been drafted [13, 14] which suggests the administrative basis for the BRMF should be changed from employer data to income tax data, the former being used as a supplement containing more up-to-date information. The BRMF would then be capable of providing a complete list of establishments - employers and non-employers - meeting the needs of all annual surveys along the same lines as the present Census of Construction, and, at the same time, better supporting a small area business program. Under this proposal, sub-annual surveys would be linked to their annual counterparts and the whole program centred on the income tax defined universe. Employer data would be sampled to extract current information.

As part and parcel of the proposal a new BRMF system would have to be designed. It would include provision for storage of relevant administrative data, and of multi-establishment structures for enterprises defined by the Intercorporate Ownership program. Implementation of the proposal would have to be incremental over a period of perhaps several years. It would be a formidable task in view of the size of the data files involved, the complexity of the procedures and the need to avoid disruption of regular survey operations.

The 1980 SIC [2] is currently being implemented. It will provide standardization of establishment definition and classification practices. A separate but related classification scheme suitable for multi-activity companies [15] is also being introduced. As part of the implementation process all SIC codes on the corporate tax file are being reviewed and 1980 codes are being assigned.

Multi-Establishments

Enterprise identifiers from the Intercorporate Ownership files will soon be put on the BRMF. This will greatly facilitate establishment profiling. A new program is being considered for verification of establishment structures, along
the lines of the Company Organization Survey at
the U.S. Bureau of the Census.

Coverage
Implementation of the new proposal previously
described would provide for comprehensive cover-
age of the non-employer as well as the employer
universe, therefore meeting the frame data
requirements of annual surveys involving substan-
tial numbers of non-employers. In the meantime,
BRMF corporate tax data linkages are being estab-
lished to facilitate use of tax data via the BRMF
and exchange of classification information
between BRMF and corporate tax files.

Systems
Generalized software is being developed for
use in conjunction with the BRMF to coordinate
all "nature of business" enquiries for frame
data. The "Company Contact System" for recording
and unduplicating contacts already exists in
embryo form [16]. It will eventually be incor-
porated into a more comprehensive vehicle for
specifying, sampling, collection, capture and
processing of frame data.

Implementation of the new proposal previously
outlined would provide an opportunity for re-
placement of some of the more restrictive fea-
tures of the current BRMF system, in particular
for revision of the present identification system
and improvement of facilities to link and store
frame data from all major sources.

CONCLUDING REMARKS
An examination [17] of counterpart programs
for the provision of economic survey frame data
in the U.S. suggest that many similar problems
are encountered there. This statement is proba-
bly equally applicable to other industrialized
countries too. Thus, potentially, there is much
to be gained by exchanging experiences - which is
the primary motivation for this paper.

REFERENCES
All references are to Statistics Canada in-
ternal working documents except where otherwise
indicated. Contact the author (613-995-9911)
regarding copies.

1. National Income and Expenditure Accounts,
   Volume 3, Statistics Canada, Catalogue
   13-549E.
   Statistics Canada, Catalogue 12-501E.
3. Policy for the Implementation and Use of the
   Business Register and for the Conduct of
   Business Profiling, Business Register Man-
4. Business Register / Manufacturing and Primary
   Industries Division Protocol. Bryan Finn et
   al., October 1981.
5. Business Register / Labour Division Protocol
   Gerry Charron et al., June 1981.
6. Merchandising and Services Divisional Master
   File/ Business Register Master File Protocol
   A. Monty, C. Bordeleau, June 1980.
7. Construction Division: The Census of Con-
   struction P. Giles, July 1983.
8. Large Scale Imputation of Survey Data M.J.
   Colledge, J.H. Johnson, R. PerS, I.G. Sande
   Proc. of Section on Survey Research Methods
   (pp431-436) American Statistical Association,
   1978
9. Estimators Based on Several Stratified Sam-
   ples With Applications to Multiple Frame Sur-
   veys. M. Bankier. To be presented at the

1983 Annual Meeting of the American Statis-
tical Association.
    M. Colledge, G. Sande, November 1980.
11. Estimates Produced from the Business Register
    Quality Assurance Study V. Estevao et al,
    July 1983.
12. Employment Statistics Development Project:
    Sample Selection and Rotation. F. Maranda,
    D. Beckstead, June 1980.
    M. Colledge, March 1983.
14. Small Area Business Data: A Proposal
15. Company Classification System (Draft). Stan-
    dards Division, December 1992.
16. Business Register Company Contact System
    Detailed Statement of Requirements.
17. Frame Data For Business Surveys: Report on
    Visit to Three U.S. Agencies. Michael
    Colledge, Andre Monty, John Pidcock, July
    1983.

APPENDIX: SOME OF THE MAJOR ECONOMIC SURVEYS
The following list, organized by subject
matter division, indicates some of the major eco-
nomic surveys at Statistics Canada for which frame
data are or could be provided from adminis-
trative sources. The sources currently used are
indicated in parentheses, "BRMF" referring to em-
ployee data, "tax data" referring to income tax
data and "ICO" to intercorporate ownership data.
The list does not include surveys for which a legisla-
tive process provides a tailor-made frame,
nor monthly surveys which depend upon the corre-
sponding annual survey for the frame.

Business Finance Division
Quarterly Surveys of Corporations (tax data)
Annual Survey of Corporations (tax data)
Corporations and Labour Unions Returns Act Survey
(annual) (tax, ICO data)
Construction Division
Census of Construction (annual) (tax data,
BRMF)
Capital Expenditures Survey (annual)
Labour Division
Survey of Employment, Payrolls and Hours
(monthly) (BRMF)
Labour Compensation Survey (occasional) (BRMF)
Manufacturing and Primary Industries Division
Census of Manufactures and Forestry (annual)
(BRMF)
Merchandising and Services Division
Retail Trade Survey (monthly) (BRMF)
New Retail Trade Annual Survey (annual) (BRMF)
Monthly Restaurants Caterers and Taverns Survey
(BRMF)
Traveller Accommodation Survey (annual) (BRMF)
Personal Household Services Survey (annual) (tax
data)
Amusement and Recreational Services Survey
(annual) (tax data)
Business Service Industries Survey (annual) (tax
data)
Transportation and Communications Division
Motor Carriers - Freight and Household Goods
Movers Survey (annual) (BRMF)
For-Hire Trucking Survey (annual) (BRMF)
Motor Carrier Passenger Surveys (annual) (BRMF)
Intercity Bus Passenger Origin and Destination
Survey (annual)
Survey of Water Transportation (annual) (BRMF)