Statistics Canada Statistique Canada

Methodology Branch

Direction de la méthodologie

Business Survey Methods Division

Division des méthodes d'enquêtes entreprises

Canada

[redacted]

#50423

2.2

STRATEGY FOR THE PROVISION OF FRAME DATA AND USE OF INCOME TAX DATA IN ECONOMIC STATISTICS

M. Colledge et all

Working Paper No. BSMD 85-045E

STRATEGY FOR THE PROVISION OF FRAME DATA AND USE OF INCOME TAX DATA IN ECONOMIC STATISTICS

Final draft July 5, 1984 Anna Ansmits, Michael Colledge,
André Monty, Innis Sande, Alf Winkworth.

1. INTRODUCTION

1.1 Motivation and Terms of Reference

The reasons for developing a general "Infrastructure Strategy" for the provision of frame data and use of income tax data are well known:

(a) it should lead to improvement in data quality as a result of standardization, better coverage, more coherence of procedures and consistency of data;

(b) it should lead to improvement in efficiency as a result of coordination of activities, including provision of centralized services where these can be demonstrably justified on a cost-benefit basis.

A Task Force to develop an Infrastructure Strategy was established by the Deputy Chief Statistician, Dr. I.P. Fellegi in March, 1984. Its terms of reference were, in brief:

(a) to define in detailed terms a strategy for redeveloping frame systems and data to service the programme of economic surveys.

(b) to identify the elements of such a strategy which are perceived to be of immediate benefit and for which implementation could begin in 1984/85;

(c) to identify the elements of such a strategy which may cause problems, by being difficult to implement;

(d) to define procedures for analyzing such problems and for testing possible solutions.

This report has been prepared with a dual purpose in mind. Firstly, it is the primary vehicle by which the Task Force is fulfilling its terms of reference. Secondly, it can serve as a draft strategy for dissemination and discussion throughout the Business Statistics Field, hence providing the basis for development of a more authoritative document.

In the short time-frame allowed for development of the document it has not been possible to contact all the persons involved in the provision of frame data and use of income tax data. It should be understood that their input is vital to the production of a definitive Infrastructure Strategy.

1.2 Scope of the Strategy

The Bureau requires an overall "Strategy for Economic Statistics" which identifies the global demands for data and which specifies, first, the data to be produced, including the required levels of aggregation and quality, and, secondly, the production methods. This report may be regarded as providing one component, covering all aspects of the provision of frame data and use of tax data for economic surveys, of such an "Overall Strategy".

Administrative records are the major source of frame information. They provide data which are not only useful for frame construction and maintenance purposes but which can supplement or replace survey data. This is particularly true of Revenue Canada income tax returns. As the uses of such data for "frame" and for "survey" purposes are interrelated this strategy includes both aspects.

In concept a strategy for frame data should cover all surveys of economic statistics, including those in the System of National Accounts, Agriculture Division, Health and Science Division, etc. However, this document focusses attention more or less exclusively upon the activities in Business Finance, Construction, Labour, MAPID, Merchandising, Transportation and Business Register/Tax Record Access Divisions. There is no reference to the Farm Register. This substantial deficiency in the strategy should be acknowledged, and rectified in due course.

It was impossible to develop an approach to the provision of frame data without assuming some overall programme for economic surveys, i.e., a framework within which frame requirements could be identified and the constraints on satisfying them determined. The Bureau programme which was assumed as a starting point is outlined in Section 2. As it provides the basis for the strategy developed, any significant omissions or errors it contains should be identified as soon as possible. In particular, the assumed nature and extent of the small area business data programme have to be validated as this programme could profoundly influence the procedures for acquisition of tax data and the consequent provision of frame information based on such data.

At the outset it should perhaps be stated that the Infrastructure Strategy developed in this document does not embody any really dramatic departures from current practice, rather it is an attempt to rationalize, co-ordinate and improve present procedures and systems. Some of

the individual tasks involved in validating and implementing the strategy are already in progress; none appear to be unmanageable. In total however these tasks will constitute a considerable amount of work touching virtually all elements of the economic statistics programme.

1.3 Contents

A description of Bureau programme, and the corresponding requirements for frame data and income tax data and current procedures for satisfying them are summarized in Chapters 2, 3 and 4. The Infrastructure Strategy itself is presented in Chapter 5 where the target systems and procedures are defined, accompanied by notes regarding their validation and implementation. These implementation notes are regrouped and elaborated in Chapter 6 to form a general implementation plan. Chapter 7 details the immediate tasks and short term benefits associated with such a plan.

2. BUREAU PROGRAMME FOR ECONOMIC STATISTICS

As stated in the Introduction, a prerequisite for development of a frame strategy is an explicit statement of the overall Bureau economic statistics programme for the next five years or so, i.e., a list of the major surveys and their separate and collective objectives. In the absence of a definitive document, the task force made certain assumptions about the likely content of such a document. These assumptions are presented in the following paragraphs. It should be noted that they do not necessarily reflect the opinions of the task force as to what should be in the programme (which would require another study), but rather what is likely to be there. It must also be reiterated that, as the programme outline in the following paragraphs is the basis to the strategy, it should be carefully validated and the significance of any omissions or errors assessed.

The Business programme is reviewed here strictly from the perspective of centralized provision of frames. An attempt has been made to group together all programme elements which are in some sense homogeneous in this respect, and to identify separately elements or groups of elements which are distinct.

The major components of the Bureau programme for economic statistics are assumed to be as follows:

(a) Annual "Benchmark" Surveys of Economic Production

These annual surveys involve collection and publication of the full range of economic production statistics, at 3-4 digit (1980) SIC by province/major urban area level for all "larger" businesses, and at 1-2 digit SIC by province/major urban area for businesses below industry specific size cut-offs. The data are obtained on a statistical establishment basis. The set of establishments are divided mutually exclusively between the various surveys on the basis of SIC grouping. Collection of activity and commodity data are limited essentially to the larger businesses. Use is made of administrative data for smaller businesses.

This set of surveys includes: the Census of Construction, the Census of Manufactures and Logging; annual surveys of surface and marine transportation; Annual Wholesale Survey, Annual Retail Survey; annual surveys of services, all of which are viewed as components of a comprehensive "annual programme for economic production".

(b) Annual Employment and Labour Income Programme

This programme makes use of data derived by aggregation of monthly survey data, possibly supplemented by an annual data collection exercise, together with data obtained annually from RCT payroll deduction/income tax (T4/T4A summaries). Industrial coding is at 2-4 digit level, depending upon industry. The establishment is the basic statistical unit thereby facilitating comparison and integration of the data with the output from the annual programme for economic production.

(c) Annual Programme for Capital Wealth and Expenditures

This programme has two components: a survey of annual capital expenditures (forecast, preliminary and actual), and a system for updating capital stock estimates based upon expenditures. Capital expenditure intentions and data are not always available from establishments thus indicating the need for a higher level unit ("capital cost centre"). The data are published at 3-4 digit SIC by province level.

(d) Annual Survey of Corporations Financial Statements

This programme is based more or less exclusively upon data obtained from corporate (T2) income tax returns. The basic statistical unit is the corporate tax-filer. Data are published at 2-3 digit SIC level and may in the future be broken down by province.

(e) Annual CALURA/Intercorporate Ownership/Balance of Payment/Multi-National Enterprises Programme

The Act (CALURA) defines a set of statistical units termed "enterprises" based upon ownership and control considerations. This set of units can, with relatively minor modifications, be used for some Balance of Payments surveys and for the Multinational Enterprises Programme. Data are published at 2-3 digit SIC level.

(f) Annual Programme for Small Business Statistics

This programme produces income, outlay and performance estimates for small businesses (where "small" has yet to be defined) at 2 digit SIC by province level. The input data are obtained primarily from income tax files. The statistical unit is the establishment.

(g) Annual Programme for Small Area Statistics

This programme will generate income and outlay estimates for small area by industry cells (to be specified). It will be based primarily upon income tax data. The statistical unit will be the establishment.

(h) Current Surveys of Selected Production/Service Variables

These current surveys involve collection and publication of a small range of production/service variables at 2-3 digit (1980) SIC level by province/major urban area. The data are obtained on a statistical establishment basis, but the target population may be defined in terms of activity not industry. The frame data must be up-to-date in order to reflect current trends. The data are expected to

project, in some sense, corresponding annual totals into the current period. It is not anticipated that they will aggregate precisely to annual totals in view of differences of coverage.

This set of surveys includes:

Current Shipments Inventories and Orders;

Monthly/Quarterly Commodities Surveys;

Monthly Transportation Surveys;

Monthly Retail/Wholesale Trade Survey;

Monthly Traveller Acommodation Survey;

Monthly Restaurant Caterers and Taverns Survey.

(i) Monthly Survey of Employment, Payroll and Hours

This survey produces data at 2-3 digit SIC by province level and at 3-4 digit SIC, national level. The statistical unit is the establishment. The frame data must be current.

(j) Quarterly Survey of Corporations

This survey provides the quarterly complement to the Annual Survey of Corporations. As quarterly data are not available at corporate level for some corporations the statistical unit is the corporation or a grouping of corporations termed a "consolidation".

The above programme list is by no means complete. It explicitly excludes first, elements for which centralized provision of frame data is not an issue and, secondly, all elements not directly related to survey activity, i.e. service functions such as standards, methodology, etc. It also excludes, and this is a shortcoming, programmes covering Institutions, Agriculture and the SNA.

3. PROVISION OF FRAME DATA: REQUIREMENTS AND CURRENT PRACTICES

3.1. Conceptual Requirements

Each individual survey operation has need of a set of statistical units which are mutually exclusive and which provide complete coverage of the target population. For each unit in the list the following sorts of data items are required:

a) identification (number, name);

b) contact (name, address, telephone, instructions);

c) activity status (indicators: active/dormant/inactive; retail outlet/not; etc.);

d) respondent status (indicators: in-scope, in-sample for which surveys);

e) classification (industry, geographic code, size, etc., for stratification and estimation);

f) validation/auxiliary (for crosschecking, editing, imputation, estimation);

g) linkage (to other lists and longitudinal);

h) maintenance (sources, dates, quality of other data items);

i) reference period(s) (to which data are applicable).

This list of units and associated data items is termed the survey "frame".

As it is desirable to be able to combine or compare micro data from different surveys, ideally, every survey should use the same frame (or a subset of it). However, because of the differing types of data collected by surveys, e.g. production, financial, ownership, etc., and because of the different reference periods, e.g. month, quarter, year,

a single list of statistical units can not serve all purposes. There is a requirement for several distinct lists each of which satisfies the requirements of a particular survey or related group of surveys. The aim is to have as few such distinct lists as possible; also for the units in different lists to be linked and to be aggregable to one another at the lowest feasible level, the units being identical for the vast majority of businesses which are, in some sense, small. The most widely used statistical unit is (or should be) the establishment as defined in the (1980) standard industrial classification.

In addition to a set of statistical units and associated data items it is sometimes necessary to include special reporting arrangements to facilitate survey data collection from "reporting units" and subsequent allocation to the statistical units.

The basic starting points for building lists of statistical units and for subsequent frame maintenance are administrative data files. Given such files, the exercise of delineating statistical units is termed "profiling". The process of following such units through time is termed "tracking". Tracking is required to ensure consistent coverage and to facilitate longitudinal studies.

A dedicated service function should provide all the frame data and services which can more efficiently be handled centrally than by individual survey operations. In particular this includes access to and processing of most administrative files.

The conceptual requirements for frame data are expanded in a number of documents [1,2,3,4,5].

3.2 Actual Requirements

Review of the Bureau programme suggests the need for at least three distinct types of statistical unit, namely:

a) establishments - for annual surveys of production;

b) corporations - for annual and quarterly surveys of corporations;

c) enterprises (defined by ownership and/or control) - for CALURA/ICO, also Multinational Enterprises and Balance of Payments surveys.

Based on establishments, the hypothetical "annual survey of economic production" can be divided, by industry, into a set of mutually exclusive surveys, i.e. construction, manufacturing, transportation, etc.

The set of establishments supplemented by appropriate sets of reporting units should also provide the basic frame for the annual labour programme and for monthly surveys of production and employment variables. However a different set of statistical units ("capital cost centres") may also be necessary for the collection of capital expenditure data. It is also desirable to record the types of activity associated with each establishment therebye facilitating surveys of activity, e.g. retail trade.

Table 1 at the end of this Chapter summarizes the statistical units associated with the major components of the Bureau programme.

3.3 Current Systems

The shortcomings of the current system for provision of frame data have been well documented, see [2], [3].

(a) the current "system" comprises a number of semi-autonomous parts, the Business Register, Tax Record Access, the Corporate Data Base, the Corporate SIC Coding Unit and the CALURA/ICO programme, between which there is inadequate linkage and consequent duplication of data and effort;

(b) frame data are not of universally good quality; frame data for large business enterprises are likely to be inaccurate due to lack of profiling maintenance; classification and status data for small businesses are also of poor quality and may differ from one survey frame to another; some units may be duplicated in supposedly mutually exclusive surveys, others not covered at all.

Deficiencies in the frame data are believed to be a major contributor to some of the current problems with estimates for certain surveys.

TABLE 1: STATISTICAL UNITS

(Survey type) Survey Name	Statistical Units		Notes
	Target	Current	
(annual, economic production)	1980 establ. tax-based		
Census of Construction	"	1970 establ. tax-based	"Own account" construction establ. not defined nor covered.
Census of Manufacturers	"	1970 establ. PD-based	"Psuedo" establishments defined to provide provincial/major urban area breakdown.
Annual Wholesale	"	1970 establ. tax-based	
Annual Retail	"		Not yet implemented; will be tax based.
Traveller Accommodation	"	1970 establ. PD-based	
Water Transport		1970 establ. PD-based	Register of vessels provides basis for frame.
Motor Carrier Freight	"	1970 establ. PD-based	"Own account" trucking establishments not covered.
Passenger Bus	"	1970-establ. PD-based	
(annual, capital)			
Capital Expenditure Survey	Company Statistical Unit (?)	mixture	Frame created by ad hoc combining of production survey mailing lists; not complete nor current.
Capital Stock	same as Capital Expenditure (?)	same as Capital Expenditure	No direct survey; estimates based on capital expenditure data
(annual, labour)			
Annual SEPH	1980 establ. PD-based	1970 establ. PD-based	Not yet fully developed; will use aggregate of SEPH (monthly) frames

(continued overleaf)

TABLE 1: STATISTICAL UNITS - Continued

(Survey type) Survey Name	Statistical Units		Notes
	Target	Current	
(annual, financial)			
Annual Survey of Corporations	Company Statistical Unit	T2 corporate tax filers	Some additions and deletions from universe of tax filers are made.
(annual, CALURA)			
Intercorporate ownership	enterprise (ICO)	enterprise (ICO)	"Enterprise" is defined in terms of ownership and control (CALURA 1983); reporting arrangements are defined by respondent and/or negociated.
(annual, other)			
Balance of Payments	company statistical unit	corporation (?)	
Multinational Enterprises	enterprise (ICO)	enterprise (ICO modified)	
(current, economic production)	1980 establ. tax-based + ?		
Current Shipments, Inventories, Orders	"	1970 establ. PD-based	
Commodity Surveys	"	1970 establ. PD-based	
Monthly Wholesale	"	1970 establ. PD-based	Being jointly redesigned.
Monthly Retail	"	Retail locations	
Restaurants, Caterers, Taverns	"	1970 establ. PD-based	
Traveller Accommodation	"	1970 establ. PD-based	

(continued overleaf)

TABLE 1: STATISTICAL UNITS - Concluded

(Survey type) Survey Name	Statistical Units		Notes
	Target	Current	
Passenger Bus	"	1970 establ. PD-based	
(current, other)			
Quarterly Survey of Corporations	1980 company statistical unit	consolidation corporation	Consolidations are groups of corporations.
Monthly SEPH	1980 establ. PD-based(?)	1970 establ. PD-based	Some division and combination of establishments to form employment reporting units

4. INCOME TAX DATA: REQUIREMENTS AND CURRENT PRACTICES

4.1. Conceptual Requirements

Income tax data are required for two interrelated purposes, first, to provide a basis for the creation and maintenance of survey frames (ref. Chapter 3), secondly, to supplement/replace survey data collection. The potential benefits of thus using tax data are the reductions in costs and in response burden.

Ideally the Bureau would like machine readable access to a comprehensive set of financial statements for every statistical unit in the universe. In practice, as the data are collected and processed by another agency, Revenue Canada Taxation (RCT), for tax administrative purposes, there are a number of limitations to their use which require identification and circumvention. In particular:

(a) there are conceptual differences between some RCT data items and those required by the Bureau, for example, RCT collect gross business income, the Bureau would prefer total shipments; the fiscal tax year does not coincide with preferred Bureau reference year;

(b) the unincorporated (T1) tax return refers to an individual tax filer who may have several distinct businesses, or may be in partnership, or both, whereas the unit of interest to the Bureau is the business;

(c) incorporated (T2) tax returns for international corporations may include data for foreign operations whereas the Bureau is concerned only with domestic economic activity;

(d) not all legal entities engaged in economic production are obliged to file tax returns, e.g. government departments;

(e) large corporations and unincorporated partnerships file as part of their tax returns a single set of financial statements which does not in general provide sufficient industrial or geographic breakdown for Bureau purposes;

(f) the quality of data collection, follow-up procedures, etc., may not match Bureau standards, particularly for data item which are of peripheral interest to RCT; not all data of concern to the Bureau are data captured by RCT;

(g) the data are for a reference period of one year; there is no subannual breakdown;

(h) the data are confined essentially to financial and taxation items; they do not refer explictly to production hence are insufficient to support individual coding to full 4 digit precision of the 1980 SIC.

Taking into account these limitations the principal uses of tax data are likely to be:

(a) to help define and maintain frame data for "small businesses";

(b) to supplement annual survey collection processes by providing basic financial data for small businesses;

(c) to provide the bulk of the information required for annual small business and small area small business programmes;

(d) to provide micro data for imputation and validation purposes and to provide control totals;

(e) to source entirely the Annual Survey of Corporations Financial and Taxation Statistics.

4.2. Actual Requirements

The actual requirements for tax data arising from the likely uses to which they can be put (ref. 4.1) are summarized in the following paragraphs.

First there is the requirement for access to the universe of T2 tax filers and T1 tax filers reporting any form of self-employed income arising from economic activity. In view of the large numbers of tax filers potentially involved (0.5 million T2, 2.0 million T1) access must be in machine readable form. The data should include all items required for frame purposes, in particular industrial code size measure, geographic location and contact information. These data are needed for frame purposes.

Secondly there is the requirement for additional financial data reported by a sample of tax filers. The sample size and the detail needed depend upon the relative significance of the tax filer's economic activity and the use to which the data are to be put. For example the sampling fractions could vary from perhaps 3% for very small corporations to 100% of the very large corporations for the Annual Survey of Corporations. A very comprehensive small area business programme would require 100% sampling. The number of financial items required may range from six basic values for small businesses to one hundred or more for the Annual Survey of Corporations.

As these data are not all captured by Revenue Canada for administrative purposes there must be procedures for sampling the tax return documents and transcribing the required data items. Typically the sample will comprise a number of components; namely:

(a) a "cross sectional" sample - spanning all industries, stratified by (current reference year values of) province and size to ensure coverage and efficiency respectively;

(b) a "longitudinal" sample - perhaps 2%, for analysis purposes;

(c) "prespecified" samples - defined independently by various user surveys to meet their special needs in so far as the cross sectional sample can not do so.

Obviously maximal cooperation between RCT and the Bureau is needed to ensure all data capture, coding sampling and transcription processes are as efficient and effective as possible.

4.3. Current Practices

The sampling and acquisition practices as of 1981 were documented very comprehensively by Valiquette and Adams in their reports [11, 12]. Apart from the elimination of the duplicate T2 return for all but the very longest corporations, and the consequent redevelopment of sampling and acquisition procedures [13] there have been no other major changes to date, although certain significant RCT initiatives are in the pipeline and the Bureau now has a mini-computer for tax data. Thus the problems documented by Valiquette and Adams remain, and there are others. In brief they may be summarized as follows:

(a) sampling, transcription and SIC coding procedures are not fully coordinated between the Bureau and RCT, nor indeed within the Bureau itself, with the result that there are duplications and discrepancies, and that less than optimal use is made of all the data transcribed;

(b) the current alternative levels of transcription detail do not meet all user survey needs;

(c) accurate up-to-date SIC coding is not maintained;

(d) the various programmes utilising the data employ different procedures with the consequent potential for inconsistent results.

5. ELEMENTS OF THE STRATEGY

Sections

A: General Policies and Principles

B: Central Service Function

C: Central Frame Data Base

D: Acquisition and Use of Income Tax Data

E: Acquisition and Use of Payroll Deduction Data

F: Acquisition and Use of Other Administrative or Commercial Data

G: Acquisition of Frame Data by Specific Direct Contact

H: Acquisition of Frame Data from Survey Routine Data Collection

I: Central Service Function: Standards, Procedures and Software

J: Annual Surveys of Economic Production

Each section is preceded by brief explanatory notes and followed by comments on its validation and implementation.

A. General Policies and Principles

Notes

The starting point for any policy or procedures should be an "Overall Strategy" defining the framework within which the objectives of all individual surveys and service functions are to be formulated. This "Overall Strategy" should indicate the balance between survey user requirements (usually for fine detail) and the level of detail which can be supported by available data sources and survey resources. The target populations, statistical units, users and publication objectives of surveys determine to a major extent the frame requirements and potential use of tax data.

The general policies and principles introduced in this section may be summarized as the rienforcement of centralized services and data base, an increased emphasis upon frame maintenance for large businesses, and the introduction of standardized methodology for annual surveys of production and the use of income tax data. These policies are further elaborated in later sections.

Strategy

A1. There will be an "Overall Strategy for Economic Statistics". It will outline the framework within which all development of economic statistics should take place.

A2. The "Strategy for Provision of Frame Data and Use of Income Tax Data in Economic Statistics", i.e. this document itself, will, after review and modifications as necessary, be confirmed as the infrastructure component of the "Overall Strategy".

A3. Provision of frame data and use of income tax data for economic surveys is a central function. It will be based on a Central Frame Data Base (CFDB) and coordinated by a Central Service Function (CSF).

A4. Units on the CFDB will be classified by size, complexity, etc. (See C11). Priority will be given to the maintenance of units which have been categorized as "Class 1" by virtue of their large size, the multiple nature of their relationships with other units or their substantial impact upon the operations or estimates of a survey.

A5. Every survey of economic statistics will be a CFDB "user", i.e. draw its frame from the CFDB, unless an alternative source and set of statistical units has been explicitly recognized within the overall strategy.

A6. The frames for all annual surveys of economic production will be derived from the same base of large establishments complemented by income tax data.

A7. There will be a "generalized methodology" for annual surveys of economic production so that, in effect, they can be viewed as elements of an all-industry annual economic survey. It will contain provision for accommodating the special requirements of each survey.

A8. Policies and procedures will be issued, based on this document, defining more precisely the CSF, CFDB and individual survey responsibilities and interfaces.

Implementation

An "Overall Strategy for Economic Statistics" should be formulated and disseminated following a systematic review of user requirements, data sources and data collection systems. In particular the business statistics survey programme components should be specified with the target populations fully defined.

The "Strategy for Provision of Frame Data and Use of Income Tax Data", i.e. this document itself, should be reviewed, modified as necessary, confirmed and disseminated as one component of the "Overall Strategy". It should ultimately be extended to include all economic surveys (e.g. of institutions, farming, governments, etc.) and to take into account future additions to the programme.

The existing BR policy document should be reviewed and replaced by a comprehensive policy outlining the target roles of the Central Service Function and user surveys. Plans for phased transition to this target from the current status should be prepared.

Each individual survey in turn should be reviewed from a Bureau perapective in terms of the "Overall Strategy". In particular, all frame-related aspects (concepts, data methodology, systems, procedures) of each aurvey should be reviewed. The appropriate statiatical units should be identified. A set of units other than 1980 SIC establishments should be explicitly justified. The

discrepancy between current practice for provision of frame data to the survey, and the target system, should be analyzed. A plan for transition should be prepared and the impact of transition assessed.

Implementation of transition plans should be scheduled on a survey-by-survey basis, preferably, but not necessarily, coinciding with survey redesign. Conversely, every redesign should incorporate a plan for transition. It should be acknowledged that the implied changes may well be profound and may ultimately cause discontinuities in the published series. Considerable effort will be required to analyze each situation and decide how it should be handled. For example preliminary validation studies may be required, followed by some type of parallel run. A new publication format may be necessary. Once the impact of transition has been assessed a certain amount of publicity outside the Bureau might be advantageous to prepare consumers for the changes and to hear their views.

A Steering Committee should be set up including CSF, divisional, SNA and survey representatives to monitor, control and facilitate implementation of this frame and tax data strategy.

B. Central Service Function (CSF)

Notes

A central function should be responsible for all those services which can be more efficiently and effectively provided centrally than by individual survey operations. In the context of this strategy two basic services can be identified, first, the provision of frame data, secondly, the sampling, acquisition, transcription, etc. of income tax data. These functions are intimately related. Frame data maintenance depends upon income tax data. Conversely frame data are required to supplement, validate and weight sampled tax data.

There are other services which must be provided centrally but which are not included in the CSF as defined in the following paragraphs, namely:

(a) development and monitoring of policies, standards and guidelines;

(b) development of statistical methodology;

(c) development of systems and procedures.

These functions are performed by staff in Standards, Methodology and Systems Divisions.

Strategy

B1. The role of the Central Service Function (CSF) will be to undertake and coordinate all the diverse functions associated with the provision of frame data and with the acquisition and use of administrative data which can be more efficiently and effectively accomplished from a corporate point of view on a centralized basis than by individual survey or divisional operations.

B2. The CSF will work in consultation with the subject matter divisions. There will be an interdivisional working group to resolve common problems.

B3. Two different but interrelated functions will be identified:

CSF(1): provision of frame data

CSF(2): sampling, acquisition, processing and weighting of income tax data.

Appropriate monitoring, quality, validation and user assistance activities will be included within each function.

B4. The CSF(1) will manage the CFDB and will provide to each CFDB user survey the full range of services associated with frame data concepts, acquisition, maintenance, quality assurance, and usage. In particular it will provide:

(a) a list of statistical units and associated frame data items for any specified survey and reference period; the list will cover every unit in the survey universe either directly or by statistical sample, i.e. a weighted subset of units which aggregate to the total;

(b) facilities for the maintenance of these data using administrative sources and using information obtained from the direct contacts made during routine survey data collection;

(c) facilities for acquisition of additional frame data by specific direct contact (profiling, self-profiling or nature of business enquiry);

(d) facilities for sampling the lists (and for mailing);

(e) standards, guidelines, procedures, software, reports and studies relating to the acquisition, processing and quality assurance of frame data.

B5. The CSF(1) will be supported by the following systems:

(a) a Central Frame Data Base (CFBD) for the storage, maintenance and dissemination of all frame data items acquired from administrative sources and/or survey operations;

(b) subsystems for the processing of all relevant administrative data files and extraction of data to build and maintain the lists of statistical units for specified reference periods;

(c) subsystems for the coordination of all specific frame data enquiries (profiling, self-profiling, nature of business) and for the data capture, processing, storage and dissemination of such data.

(d) subsystems to interface each survey or group of surveys to the CFDB.

(e) subsystems for unduplicating units and for processing, editing cross-checking and manitoring selected data items (in particular gross business income and/or salaries and wages) from various surveys.

B6. The CSF(2) will coordinate all survey sampling of income tax data. It will undertake the acquisition, transcription, editing, imputation, estimation and dissemination required for the use of such data by the various programmes.

B7. The CSF(2) will be supported by a sampled tax data base linked to the CFDB.

Implementation

Central services are presently provided by four separate groups: Tax Record Access, Business Register, Business Finance Corporate Data Base and Business Finance SIC Coding and Rebasing Unit. Transition to the target CSF will probably require organizational restructuring. This should be carefully planned. All individuals likely to be affected should be well briefed beforehand and their comments taken into account. Good morale is essential. The structure of the future CSF may include units for operations, systems, analysis, planning, quality assurance and administration.

Transition should begin with the standardization of concepts and procedures, the coordination of existing operations and dissociation of future central services from existing survey specific functions in Business Finance Division.

Transition should be phased. It should be transparent to user surveys, or should incorporate explicit provision for affected surveys to make the attendant changes.

The mini-computer system currently being developed within Tax Record Access should ultimately incorporate all sampled tax data.

C. Central Frame Data Base (CFDB)

Notes

The broad principles underlying this section are, first, the integration of various source and statistical data files, secondly, the emphasis on maintenance of the most important units and, thirdly, the replacement of PD data by income tax data as the primary basis for survey frames.

All the data envisaged as part of the Central Frame Data Base (CFDB) are currently within the Bureau but in various separate files between which linkages have been established, to a lesser or greater extent, but never systematically maintained. These files and linkages will provide the starting point for CFDB initialization.

Total integration of all data files does not seem worthwhile due to the vast numbers and volatility of small businesses. As a compromise between cost and quality it is proposed to ensure linkage only between all large or otherwise significant units. These same units will receive priority in terms of quality of maintenance.

For smaller units two alternative sets of establishments will be available, one based on income tax data, the other on PD data. The former will be preferred for all annual surveys of economic production as it will provide full coverage of both employers and non-employers and as it will facilitate use of income tax data to supplement or replace survey data. For current surveys neither will be fully satisfactory due to time lag, though the PD based frame will be more up-to-date. Use of multiple frame techniques based on tax, PD and possibly area frames are recommended. The extent to which annual and current survey estimates are discrepant due to the use of different (tax-based, PD-based) frames will be monitored. In any case, perfect consistency can not be expected for other reasons (non-additive measurements, etc.)

Strategy

(Items C1-C5: Function of CFDB)

C1. The overall function of the CFDB will be to store, maintain and provide access to all the data of current or potential use for frame purposes which are relevant to more than one survey, which are useful for validation, or which can be stored and maintained more effectively in a central repository than in survey-specific data files.

C2. The CFDB will provide universe frame data to each user survey, for any specified reference period, for either of two main purposes:
(a) to enable selection of a sample for mail-out or tax acquisition;
(b) to obtain counts and weights for estimation.

C3. For each individual survey a prescribed protocol will define CFDB access and update facilities. This protocol will include provision for direct access and update and will enable the survey to view the CFDB as an integral part of the survey front-end system. Data on the survey in-sample master file will be synchronized with CFDB data. Ideally the in-sample master file will be an extension of CFDB software. Separate survey-specific universe file systems and data files will not exist unless their need is explicitly recognized in the "Overall Strategy". (Also see Sections J, K).

C4. The CFDB will contain facilities for checking integrity. For example comparisons will be made of data on wages and salaries derived from different sources and aggregated to a common unit level. These data will be available for analytic purposes, e.g. to the SNA.

C5. The CFDB will provide facilities for monitoring units in sample and for sample selection from any specified frame by any standard technique, possibly incorporating response burden constraints if and when appropriate techniques have been developed.

(Items C6-C10: Contents of CFDB)

C6. The CFDB will be the repository for the storage and maintenance of all relevant administrative and commercial lists of units and associated data items, specifically including:

(a) corporate (T2) and individual (T1) income tax return data extracted from RCT machine readable files or by transcription from the returns;

(b) payroll deduction (PD) account data extracted from RCT machine readable files or questionnaires.

C7. The CFDB will be the repository for the storage, maintenance and access of lists of statistical units of various types, together with associated data items. The basic statistical units will be:

(a) (statistical) establishments, as defined in the 1980 SIC;

(b) legal entities, as defined by various acts;

(c) (statistical) enterprises, as defined by the intercorporate ownership programme.

C8. The CFDB will contain all profiling and linkage information relating units in different lists, or relating units within the same list through time. In particular the CFDB will contain complete profiling and linkage information for each unit on each list which is identified by its size, complexity or impact upon survey operations as being "large" or "significic

C9. The CFDB will contain provision for the introduction and maintenance of additional lists (and corresponding data and linkages) derived from administrative or commercial sources. It will contain provision for inclusion of additional sets of statistical units, and for the inclusion of approved survey-specific reporting units where such units can be more effectively stored and maintained centrally then by an independent survey system.

Lists of administrative or statistical units which may be added at some stage in the future to the CFDB include:

(a) Consumer and Corporate Affairs (new) charters, and the provincial equivalents;

(b) capital cost centers - statistical units for the capital stock/ expenditures programme;

(c) locations, i.e. sites of economic activity;

(d) statistical companies, defined in conjunction with the 1980 company classification system;

(c) an area frame.

C10. For each separate administrative or commercial source the CFDB will carry a corresponding "source universe" data file containing all relevant units and associated data items extracted from the source file. Such data will be reformatted, edited and imputed as need be, but will otherwise be stored in essentially the original source form, i.e. they will not be manipulated nor merged with data from other sources.

C11. The units in each source universe file will be assigned to one of five categories according to size, complexity, significance to user surveys, and the availability of associated classification and linkage data

Class I will contain units which are large or significant in some way, or which are linked to Class I units on the same or other CFDB files. The frame data associated with these units will be complete and will include all linkages to related units. The units may be "single", or may be "multi", i.e. involved in many-to-one linkages.

Class II will contain certain units which are not large nor significant but for which some linkage and all classification data are available. The units may be single or multi. It will act as a buffer to reduce the impact of units which are close to the Class I boundaries and thus have a tendency to flip-flop in and out of Class I. It will also provide a basis for storage of single or multi-units below the Class I boundary for which some linkages and all classification data have been acquired.

Class III will contain units which are small, which are treated as singles (though they may in fact be T1 partnerships or multiple businesses) and which are classified by industry but not linked.

Class IV will contain small single units not linked nor classified by industry.

Class V will contain very small or inactive units which are defined out-of-scope for most purposes.

C12. From Class I/II units on the PD, T1 and T2 universe files a corresponding list of Class I/II establishments will be built and maintained. This list will be supplemented from survey, profiling and

other administrative sources to provide complete, unduplicated coverage of all large or multi or non-tax-filing or otherwise significant establishments.

C13. Coverage of small, single establishments will be provided by the complementary, i.e. Class III/IV/V units, on the tax universe files. These units will be termed "tax-based establishments". In combination with the Class I/II establishments they will give complete (subject to time lag) coverage of the universe of economic production.

C14. In conjunction with the set of Class I/II establishments the lists of Class III/IV/V units on the PD universe file will provide complete (subject to time lag) coverage of the employer (PD) universe, thereby constituting an alternative frame of establishments to that based on income tax returns.

C15. The tax-based frame will be preferred for estimation purposes for all annual surveys. For current surveys it is suggested that the tax-based frame be supplemented by the PD-based frame and possibly an area frame, using multiple frame techniques. In the case of SEPH, the PD-based frame with or without an area frame may be preferred.

C16. The frame data items to be associated with each Class I/II, tax-based or PD-based establishment will be:

a) identification (number, name);

b) contact (name, address, telephone, instructions);

c) activity status (indicators: active/dormant/inactive; retail outlet/not; etc.);

d) respondent status (indicators: in-scope, in-sample for which surveys);

e) classification (industry, geographic code, size, etc., for stratification and estimation);

f) validation/auxiliary (for crosschecking, editing, imputation, estimation);

g) linkage (to other lists and longitudinal);

h) maintenance (sources, dates, quality of other data items);

i) reference period(s) (to which data are applicable).

A similar set of items will be associated with each other type of statistical unit.

C17. The sets of establishment units will be partitioned on the basis of SIC so that, for any given reference period, the frames supplied to annual surveys of economic production will be mutually exclusive.

C18. The frame provided to an annual survey for a given reference period will not in general be complete with respect to the smaller units at the time of sample selection. It will however, be complete at the time of estimation. Here "completeness" will be interpreted to mean providing coverage of the whole universe whether by a full set of units or by a statistical sample, i.e. weighted subset.

C19. The frame provided to a current survey will be as complete for the reference period as possible. Subsequently a complete frame for the period will be made available for revision (if any) of the survey estimates to take previous defficiency of coverage into account.

(Items C20-C23: Maintenance of Establishment Data)

C20. Priority in terms of both frequency and precision of maintenance will be given to Class I units. Additions to Class I units or changes to data items will be based primarily on direct contact i.e. by survey or by specific frame data enquiey, not upon automated processing of administrative data.

C21. Data from administrative sources will be used to validate Class I data or to signal the need for a frame data enquiry. They will be used to update Class II and III data automatically, subject to an update protocol.

C22. For every data item the date and source of the last update or validation will be recorded. This is essential for quality measurement and maintenance, in which context validation is as significant as update. For data items which are always updated or validated as block the date and source information need be recorded only once for the block.

C23. Linkages for Class II units will be maintained only to the extent that this can be accomplished automatically. If manual intervention would be required to resolve a problem with the linkage then the link will be flagged as uncertain or dropped and the unit transferred to a lower class.

C24. For Class I units SIC codes will be to 4-digit precision, updated every twelve months from annual surveys or by direct contact. For Class II/III with updating will be based a rotating nature of business/SIC self-verification sample together with automated and manual procedures for coding from tax returns. (See Item D15).

Implementation

As the first step in designing the CFDB a number of definitional and conceptual problems must be resolved, in particular:

(a) the definitions of "birth", "death" and "transformation" and the set of possible "states" for each statistical and administrative unit;

(b) the precise relationships between the various sets of statistical and administrative units;

(c) the unit identification numbering system.

The decisions should be incorporated in a comprehensive dictionary of frame and tax data terminology. They will form the basis for the design of storage, updating and integrity checking procedures.

All source files of existing or potential use for the provision of frame data (i.e. CORPAC, ADMIN FILE, COMSCREEN, SELF-EMPLOYED INCOME FILE, PAYDAC, etc.) should be analyzed systematically to establish the quality of the data and identify possible problems in their usage. The analysis should be coordinated, performed to a uniform standard and documented. It should include:

(a) universe snapshot tabulations;

(b) assessment of file data quality (missing units, missing data, data errors, etc.);

(c) longitudinal and between file data comparisons, and identification and resolution (where possible) of differences.

There should be a comprehensive, systematic review and extension of existing linkages between units on the administrative and structural unit files, (i.e. PD-SIN, PD-T2, T2-ICO, T2-BRID, etc. linkages). An assessment of the feasibility of maintaining linkages in each case should be established and documented prior to implementation of a CFDB design which is dependent upon such linkages. The linkage strategy should be cleared with the appropriate authorities within the Bureau.

Linkage between files should start with the largest and/or most significant units. These same units should also be given priority in validating automated links and in resolving multiple link and missing link problems.

Linkage procedures should incorporate all the data of potential value in establishing links, for example size, SIC, "old" linakge information, etc., not just name and address. (See item I). Procedures for maintenance of established links must be developed.

Linkage reports should include:

(a) the files being linked;

(b) the number of (in scope) units on each file;

(c) the percentage of links established at each stage in the linkage process;

(d) the procedures used at each stage;

(e) the quality of the linkages;

(f) the nature of the unlinked units.

There should be sufficient information arising from the analyses of source files and linkages between them to enable validation of the class I, II, III, IV, V concept and definition of the precise boundaries between these classes.

Procedures within the CFDB for sample selection and for monitoring sampling and response burden must be developed.

The possibility of supplementing CFDB (list based) frames with an area frame should be investigated.

CFDB systems analysis and definition of requirements could start now. It might be wise to build a small scale prototype. A period of "parallel run" with current BR/TRA systems will probably be necessary and should be planned.

A CFDB interface protocol must be established with each user survey. It must include the mechanism for transition from the BR/TRA-survey interfaces which currently exist. (Also see Items J, K).

Confidentiality issues associated with the storage of tax based data in the CFDB have to be addressed. Presently the BR is "confidential" whereas the ADMIN FIle is "secure".

D. Acquisition and Use of Income Tax Data

Notes

Income tax data are available in essentially two forms:

(a) tax return (or facsimile), including associated schedules and financial statements, from which all the data actually reported can be transcribed; the quality of such data are determined by the quality of the transcription process (which can be controlled) and the quality of reporting (which can not);

(b) RCT machine readable files, containing a subset of the reported data items, at no transcription cost, but with no explicit means of controlling quality.

This section refers to procedures for handling both forms.

The relevant RCT universe files for incorporated (T2) and unincorporated (T2) tax returns are the CORPAC file and the Self-Employed Income File (SEIF) respectively. The CORPAC file is currently processed by the Business Finance ADMIN FILE system. This processing includes editing, also incorporation of information obtained from Business Finance transcription and SIC coding procedures. Although these procedures should be evaluated, in the first instance, the ADMIN FILE might be used to advantage in place of the CORPAC file as the basic T2 universe file.

Income tax data are used for two purposes, first the provision of frame data, secondly to supplement survey data. These purposes are interrelated; both are covered in this section.

The strategy as defined is essentially a rationalization, streamlining and enhancement of the present procedures for acquisition and use of income tax data. The proposed "Weighted T1 Sample File" corresponds to the current "Weighted Combined Master File". (There is presently no T2 equivalent). The proposed "Cumulative Sample File" corresponds for T1 returns to the Historical Summary File and for T2 returns to the Admin. File. This rationalization and enhancement is necessary in view of the pivotal role which income tax data will play in the provision of frames.

Some of the returns which are sampled by STC for transcription of tax data, particularly those required by the Annual Survey of Corporations will refer to Class I/II type units. The majority, however, will refer to Class III/IV units, i.e. units which are classified but not linked or which are neither classified nor linked.

Implementation of a very comprehensive small area business statistic programme would have a profound effect upon tax data acquisition and usage. It would require that all units above some minimum size criterion were classified, implying, first, a considerable increase in resources (at RCT, STC, or both) and, secondly, improved frame data.

Strategy

D1. The identification of tax returns considered in scope for economic surveys will be determined in accordance with the target populations and objectives of economic surveys as defined in an "Overall Strategy". In particular, lower size boundary cut-offs and the inclusion/exclusion of professional and commission agent unincorporated tax filers will be specified.

D2. All relevant frame data items for each in-scope unit on the RCT universe tax files (CORPAC, SEIF) will be extracted and stored in the CFDB T1 and T2 "tax universe" files.

D3. All tax returns above a specified size threshold on the universe files will be categorized as "Class I" and will thus be linked to all corresponding units on the other major administrative and statistical lists. Additional tax returns may also be classified as Class I by virtue of their linkage to Class I units on other lists.

D4. For unincorporated class I tax returns, any partnerships or multiple businesses will be identified and the corresponding data maintained. Thus a separate statistical unit will exist for each business of the multiple. Each tax return in a partnership will be linked to the corresponding statistical unit (or units, if the partnership is a multiple business too).

D5. STC sampling of income tax returns for transcription of data will be coordinated by the CSF(2) so as to optimize in some sense, the transcription and processing resources available. The sample will include the folowing components:

(a) a "cross-sectional" (all industry) sample, selected according to CSF(2) specifications, based on the current reference year;

(b) a "longitudinal" sample;

(c) samples individually "prespecified" by user surveys, based on frame data for earlier reference periods.

D6. There will be facilities for transcription at several levels of detail in accordance with user survey requirements. In particular there will be levels which coincide with or replace those currently existing, i.e. TRA basic financial statistics, Census of Construction subsample, Annual Survey of Corporations Statistics.

D7. Estimates of counts and other totals for a given reference year in various strata will be based upon multiple frame techniques using the reference year tax universe files to provide benchmark counts. Each sampled return will be assigned a corresponding weight such that the weights aggregated over all returns sum to the universe benchmarks. These weights will be available at about the time annual surveys normally prepare their estimates.

D8. Data from all tax returns sampled in the given reference year, which are relevant to the provision of survey frames or the validation of survey data, will be linked to the corresponding units in the universe tax files and recorded in two CFDB files:

(a) the unweighted "Cumulative Sample File" containing all returns sampled in the current and previous reference years;

(b) the reference year "Weighted Sample File" containing all sampled returns and their weights.

D9. The Cumulative Sample File will provide an (incomplete) frame for survey sampling and prespecification of tax returns to be acquired for subsequent reference periods (akin to the present Historical Summary File.)

D10. The Weighted Sample File (akin to the present Combined Master File) will provide a basis for annual survey estimation, for the given reference period, for that portion of the survey universe not covered by survey (large unit) questionnaire. The data will also be useful for validation or imputation of survey data at unit level, for the small business programme and, possibly, the small area business programme.

D11. Sampled T1 returns not in Class I which refer to partnerships will be assigned a "partnership weight" on the weighted sample file. No attempt will be made to establish linkages to the other partners. The weight will compensate on a statistical basis for the fact that the data refer to a business which will also be reported on each partner's tax return, i.e. the business is represented more than once in the universe.

D12. Distinct units will be created on the weighted sample file for each separate business identified on a T1 return, reporting data for more than one business. Corresponding distinct units will be created on the unweighted cumulative sample file but will be maintained only for the reference year subsequent to which they will be collapsed to a single unit. Tax returns indicating both partnerships and multiple businesses will be handled using partnerships weights and multiple units.

D13. A policy will be developed regarding the use of other RCT machine readable income tax files (COMSCREEN, CORPAC, GREENBOOK, etc) to supplement the universe frame files in providing frame data and/or to supplement the STC transcription survey-type data, and/or for validation purposes.

D14. A policy will be developed for efficient and effective coordination of RCT and Bureau sampling, transcription and data capture processes. Changes in current RCT operations which would improve the availability and/or quality of tax data will be identified, analyzed on a cost-benefit basis and discussed with RCT. Communication with RCT will be maintained on a regular basis.

D15. Nature of business information and change of business indicators on tax returns will provide the basis for SIC code maintenance using automated coding procedures (see Item I), manual coding and nature of business/SIC self-verification enquiries. Essentially the same treatment will be given to T1 and T2 tax returns of comparable size. SIC codes for units below specified size boundaries will not be maintained on a universe basis.

Implementation

The definition of tax returns in scope for economic surveys must be determined. e.g. to include/exclude professionals, to have or not a lower size cut-off. This will require analysis of the relevant frequency distributions and the relative contributions of the smaller units to the totals at provincial and 2-3 digit SIC levels. Since the universe is very large and volatile the definition should be kept simple.

All machine readable tax files should be documented and analyzed systematically to establish the quality of the data and their potential usage for frame purposes and to supplement survey data. The CORPAC file may be of use in identifying new corporations which have not filed a return. (Also see Section C, Implementation).

An appropriate size criterion for definition of "Class I" tax returns must be selected. It should coincide as closely as possible with the corresponding PD account size criterion. It can be based only upon data items available on the T1, T2 universe tax files. Selection will require analysis of the data from linked PD accounts and tax returns. (See Section C, Implementation).

Current sampling strategies should be reviewed. The T2 cross-sectional sample scheduled for TY 84 data should be fully supported with transcription resources. The T1 cross-sectional sample design should be improved to include provision for more size strata, for sampling returns indicating any form of self-employed income and for stratification by province. Sample sizes and allocations should be rationalized. Consideration should be given to extending both designs to permit sampling by small area defined in terms of postal code and/or RCT locality code. Implementation of such improvements for TY84

Current transcription procedures should be reviewed with reference to user survey needs. Account must be taken of any new requirements arising from the small area and small business programmes.

A system for production of a weighted T2 sample file equivalent to the weighted T1 sample file should be introduced.

Data transcribed for the Annual Survey of Corporations should be made available via the CFDB to other user surveys. Conversely data transcribed for other purposes should be made available to the Annual Survey of Corporations.

A system for computing the sampling variability of estimates based on tax samples should be introduced. Without such a system comparison of alternative sampling strategies is difficult.

An analysis of the incidence and impact of T1 partnerships and multiple businesses must be made, together with an assessment of the performance of the strategy proposed for handling them.

A study should be made of the relative merits of the Combined Master (Weighted Sample) File versus the Historical Summary (Accumulated Sample) File as the basic frame file for T1 estimates. The proposed stragegy for use of the Weighted Sample File should be validated or altered according to the results of this study.

Tax data concepts and terminology should be documented and contrasted with survey data concepts. Comparisons of tax and survey data at unit level should be made to facilitate development of a strategy for handling differences. Consideration should be given to defining the annual survey reference year to coincide with the tax reference year.

Analysis of RCT machine readable files and comparison with Bureau sample files should be made with a view to making more use of such data.

Current RCT and Bureau sampling, transcription and data capture processes should be reviewed (starting with the Adams-Valiquette study documentation) with the objective of identifying operational changes which could improve the quality and/or coverage of tax data and/or reduce costs. It is evident there would be mutual gains from closer collaboration between RCT COMSCREEN, Assessing and Bureau T1 acquisition operations and between RCT CORPAC, Tax Model and Bureau T2 acquisition operations.

Changes in Bureau/RCT operations which would be beneficial should be discussed with RCT. They may include:

(a) coordination of sampling, transcription and data capture;

(b) sharing of sampled data;

(c) RCT data capture of additional items, such as nature of business information, on a universe basis;

(d) RCT changes to the tax returns, e.g. introduction of fixed format schedules, prescribed lists of industrial activities;

(e) RCT coordination of PD and tax systems and linkage of identifiers;

(f) introduction of more flexible sampling facilities.

E. Acquisition and Use of Payroll Deduction (PD) Data

Notes

Payroll Deduction data are provided by RCT in four forms:

(a) PAYDAC file: machine readable file updated daily by RCT; available on request (presently acquired monthly); contains current payroll deduction data for all PD accounts;

(b) PD20 questionnaires: mailed by RCT to a new account holder shortly after the account is opened; one copy of completed questionnaire is forwarded to STC where all data including SIN (for linkage to tax files) are captured;

(c) PD-SIN pairs file: machine readable file containing all PD account holder SINs recorded by RCT; produced annually (?);

(d) T4-T4A files: machine readable files containing payroll and payroll deduction information for all employees for each PD account; produced annually.

PD10 data can be useful in signalling and classifying potentially new establishments or in identifying linkages. PAYDAC data can be useful for updating contact name and address and for signalling substantial changes in size or cessation of employer activity. PD-SIN data can be useful for creating PD-tax data linkages. (There is no equivalent PD-T2 data file). T4-T4A data can be useful in providing labour income control totals and in updating size measures. The strategy presented in this section is, in essence, a rationalization and streamlining of present processing procedures, but with the important distinction that PD data will play a subordinate role to income tax data in the construction of survey frames (at least for annual surveys of production).

Strategy

E1. All relevant frame data for each PD account on the RCT PAYDAC file will be extracted periodically and stored in the CFDB "PD universe file".

E2. All PD accounts above a specified size threshold on the PD universe file will be maintained in Class I and thus will be linked to all corresponding units on the other major administrative and statistical units. Additional PD accounts will also be maintained in Class I by virtue of their linkage to one or more Class I units.

E3. PD accounts below the specified size threshold which are involved in any form of multiple relationship, i.e. linked to one another or to more than one unit on another list, or which have been linked to a tax return, will be maintained in Class II.

E4. All data on the PD20 questionnaire will be captured, edited and passed through automated procedures for classification, and for matching and possible linkages to Class I and II establishments and PD accounts. The resulting output will be validated or corrected by CSF direct contact, prior to CFDB update, for all PD accounts above a specified size threshold.

E5. CFDB updating from PD20 data for a PD account may result in:

(a) creation of a new Class I or Class II establishment unit with corresponding frame data, and linkage to the PD account, and possibly to a tax return; or

(b) linkage between an existing Class I or II establishment and the PD account; or

(c) addition of frame data to a unit already on the PD universe file thereby creating a Class III (i.e. classified) unit.

E6. PD derived additions and updates will be of most utility in supplementing and updating income tax-based frames for current surveys. A standard, preferred multiple frame estimation procedure for handling data from tax and PD based frames will be developed.

E7. PD derived additions and updates will also be useful for supplementing the income tax-based frames used by annual surveys for sampling/mailing purposes but not for estimation.

E8. A policy will be developed for effective coordination of RCT and Bureau data collection, capture and sharing of PD information. Changes in the PD20 questionnaire which would improve the quality and/or content of the data collected will be identified and discussed with RCT. (See also D14).

E9. Procedures will be developed for making maximum use of PAYDAC and T4-T4A data for updating purposes (in particular for size coding and indication of zero employee businesses).

Implementation

An appropriate size criteria for the definition of "Class I" PD accounts must be selected which coincides as closely as possible with the corresponding tax return size criteria. (See Sections C, D, Implementation). It could comprise three distinct threshold boundaries used in hierarchical combination:

(1) a boundary in terms only of data items appearing on the PD universe file;

(2) a boundary in terms of PD universe and/or PD20 data items;

(3) a boundary in terms of PD/T4 data items.

The current Bureau/RCT interface should be reviewed and revised if need be. The issues to be addressed include frequency of access to PAYDAC machine readable files, improvements to the PD20 questionnaire and coordination of RCT and Bureau PD20 follow-up exercises.

Investigation of procedures for improved use of PD data for births, deaths and structural changes should continue, together with the analysis and reduction of time intervals involved in processing these data. Current procedures should be extended to make use of PAYDAC and/or T4/T4A for size coding, in particular to identify units which have grown into the Class I category.

F. Acquisition and Use of Other Administrative and Commercial Data

Notes

Income tax data do not provide complete coverage of the universe of economic activity. Examples of exclusions are: federal and provincial Government, charities, hospitals, schools, other institutions, etc.. Federal and Provincial Accounts are the major source of frame data pertaining to Federal and Provincial Governments.

Income tax data do not provide current coverage of economic activity. Possible sources of more up-to-date information, apart from PD data, are:

(a) Federal and Provincial lists of corporate charters;

(b) municipal lists of businesses;

(c) hydro-electricity accounts, telephone accounts, etc.

The difficulty in utilizing data from such lists is invariably the absence of a common identifier and the consequent problem of duplication.

Certain legislative and administrative processes generate tailor-made frames for specific survey purposes, e.g. airline regulations generate a frame for air transport. The use of such frames should be explicitly accepted in the "Overall Strategy".

A common identifier for all businesses/economic operations would greatly facilitate linkage of data from diverse sources.

Strategy

F1. Commercial data sources and administrative sources other than RCT T1, T2 and PD systems will be reviewed periodically by the CSF to evaluate their potential utility for improving coverage and currency. Frame data will be extracted and incorporated in the CFDB where appropriate.

F2. In the absence of identifiers providing exact matching between these lists and the CFDB, the use of multiple frame estimation techniques [6]-[10] will be investigated as a possible mechanism for incorporating data for samples from these lists.

F3. The Bureau will seize every opportunity to promote the concept of a single, all-purpose identifier for every corporation and for as many unincorporated businesses/economic operations as could reasonably be covered.

F4. A survey which derives its frame from the CFDB may, in collaboration with the CSF, make limited use of other administrative or commercial lists for validation. (This is not a reference to frame data enquiries by direct contact which will be coordinated by the CSF). Any data thereby obtained which refer to items on the CFDB should be transmitted to the CFDB, and this should be the only mechanism by which such data update the survey frame.

F5. Certain surveys, as specified in the "Overall Strategy", will have target populations for which coverage is provided by non-standard, survey-specific set of statistical units obtainable from commercial or administrative sources. In such cases survey operations will maintain the survey frame directly from such sources not via the CFDB. Any data thus obtained of relevance to the CFDB will be transmitted to it by survey operations. Linkage with the CFDB may be attempted or maintained, however, for validation purposes.

Implementation

The list of surveys which will maintain their frames directly and not via the CFDB should be established as part of the "Overall Strategy (ref. Item A). See Items C, J, K (Implementation) for related notes.

France provides an example where businesses are legally registered and assigned an identifier by the statistical agency.

G. Acquisition of Frame Data by Direct Contact

Notes

"Direct contact" here implies contact with the entity in question, by telephone/mail questionnaire etc., for the explicit purpose of obtaining frame data. It thus excludes acquisition of data from either administrative sources (ref. items D, E, F) or as part of a normal survey data collection operation (ref. item H).

Procedures for obtaining data by direct contact may be divided into three main groups:

(a) profiling - for large and/or complex entities; involving preparatory investigation of the possible profiles, personal interview with representatives of the entity, delineation of statistical units and negociation of reporting arrangements;

(b) self-profiling - for medium size/complexity entities; involving a request for validation or correction of a profile by mail;

(c) nature of business enquiry - for "single" units, involving mail or telephone enquiry to obtain frame data.

It has been current practice for survey operations to undertake enquiries of various kinds independently of one another and of the Business Register. This has resulted in duplication of effort, unnecessary response burden and lack of data exchange. Thus there is the need for a coordinated programme of enquiries and updates. As resources are limited priorities and allocations to the various surveys will have to be made.

Strategy

G1. The circumstances which can trigger acquisition of frame data by direct contact will be prescribed.

G2. The procedures for acquisition, processing, monitoring and dissemination of data by direct contact will be standardized and carried out under the control of the CSF using the Company Contact System (CCS) for processing and control and the CFDB for storage and dissemination.

(Items G3-G4: Nature of Business Enquiries)

G3. Individual survey operations may request nature of business enquiries, subject to prescribed rules and survey budgetary constraints. The enquiries will be carried out and processed by the CSF using standardized procedures.

G4. The CSF will itself initiate nature of business enquiries:

(a) to resolve problems arising during the course of processing administrative data;

(b) to resolve problems referred to it by survey operations;

(c) as part of CFDB quality assurance.

(Items G5-G8: Profiling)

G5. Profiling of very large and/or complex entities should be carried out on the basis of an established profiling interview plan. The plan should contain provision for resolution of ad hoc survey problems as they arise, in addition to a routine maintenance schedule.

G6. Interviews will be carried out according to defined profiling procedures.

G7. All actual or potential Class I multi-establishment enterprises not in scope for a profiling interview will be covered by a self-profiling survey.

G8. All profiling will be coordinated with procedures for maintenance of CALURA/ICO enterprise structures and with annual surveys of economic production. Profiling data will be stored in a CFDB subsystem.

Implementation

Recommended profiling procedures and plans should disseminated for general comment, revised as need be and implemented.

A self-profiling survey must be developed. Existing methods used by the U.S. Census Bureau and, jointly, by the Quarterly Survey of Corporations and the Capital Expenditures Survey may provide some guidance.

Present procedures should be rationalized and standardized to use as few different forms as possible. The Company Contact Syatem will accept any statistical or administrative identifier and will unduplicate requests making use of CFDB linkages as necessary.

Criteria for nature of business enquiries and budgets for each survey must be established.

H. Acquisition of Frame Data During Survey Data Collection

Notes

"Survey data collection" here refers to the collection of data by routine survey process, i.e. mail questionnaire or other collection vehicle. It does not refer to frame data enquiries, even though these may be initiated during the course of survey operations. (See item G).

The main thrust of this section is that surveys should use standard concepts and procedures whenever possible and, in particular, that all frame data items even those collected by survey operations should be channelled through the CFDB. This will help to obviate the need for survey-specific universe frame files and to promote frame data exchange.

Strategy

H1. All frame data obtained during the course of survey data collection which refer to items on the CFDB will be transmitted to the CFDB. This will be the only mechanism by which such data items can be used to update the survey frame. (Ref. Item C3).

H2. The questionnaires for annual surveys of production will request sufficient data to validate/correct all frame data items for the statistical units (establishments) in sample. Such data will usually be regarded as definitive for the given reference period and will cause appropriate updating of the CFDB.

H3. Frame data derived from other types of survey (than annual production) and transmitted to the CFDB will be processed in accordance with a CFDB updating protocol. (Ref. Item C3) This protocol will comprise a set of rules to resolve data conflicts on the basis of the source and reference period to which the data refer. Depending upon their nature, survey based updates may be applied directly to CFDB data items or may require CSF validation.

H4. All surveys will use standardized concepts, definitions and procedures except where specific exceptions are recognized in the "Overall Strategy". Survey questionnaires will be evaluated for consistency with the concepts and one another.

Implementation

Annual production survey questionnaires must be reviewed and revised if need be to include requests for information to validate or update all frame data items. In particular it must be ensured that sufficient data are collected to support full precision, 4 digit SIC coding and to detect changes in establishment coverage.

The CFDB must be designed to provide user surveys with the speed of update turn-around they require for universe frame maintenance. Also see Sections C, I, J Implementation for related comments.

I. Central Service Function: Standards, Procedures and Software

Notes

This Section refers to standards, procedures and software which must be in place to support the policies outlined in previous Sections.

Presently there is a standard [1] defining the establishment statistical unit and SIC code but it has not been fully implemented. A "Company Classification System" [14] has been formulated but not yet formally adopted. Procedures for actual assignment of SIC codes are not completely standardized. There are a variety of different geographical coding schemes in current use, creating problems of consistency and confidentiality. The purpose of this section is to promote standardization and automation.

Strategy

(Items I1-I2: Standards and Procedures)

I1. There will be standards for industrial and geographic coding for all types of statistical units.

I2. There will be procedures and plans for implementing the standards.

(Items I3-I10: Software)

I3. There will be a Company Contact System to facilitate and coordinate all procedures for the acquisition, processing and dissemination of frame data enquiries.

I4. There will be software for automated assignment of SIC codes to nature of business descriptions, primarily for use in coding smaller units.

I5. There will be software for automated assignment of SGC given postal code, and of postal code given address.

I6. There will be a self-profiling survey system.

I7. There will be quality assurance systems.

I8. There will be software for automated linkage which can utilize all pertinent data items in combination, according to their value in establishing links.

I9. There will be software for frame data report generation.

I10. There will be software to interface the CFDB to user survey systems.

Implementation

Standardized procedures for SIC coding, including assignment of "degree of certainty" indicators, or equivalent, should be introduced.

The plans for conversion of existing establishment units to 1980 SIC codes should be monitored and revised if need be.

A plan and procedures for introduction of 1980 establishment statistical and reporting units must be prepared and implemented in harmony with the profiling programme (see Item G). Similarly plans and procedures for introduction of the 1980 company classification system and company statistical unit should be prepared.

Standardized geographic classification and coding procedures should be implemented.

The existing Business Register, Tax Record Access and Business Finance quality assurance functions should be integrated and adapted to the future CFDB.

Software for report generation and user survey interface should be designed and developed with the new CFDB system (see Section C). The self-profiling survey system, Item I6, has to be developed from scratch. The remaining systems should all be obtainable with relatively minor modifications from software which exists now in prototype or production form in the Business Register and elsewhere. For example enhanced linkage procedures may be developed using GIRLS [15] in conjunction with existing BR routines.

J. Annual Surveys of Economic Production

Notes

The underlying objective of this section is to provide a consistent procedure for all annual surveys of economic production so that they can be collectively regarded as constituting a coherent "annual economic survey".

A common frame, sampling and estimation procedure applied by all annual surveys of economic production is the only effective means of ensuring that coverage of each individual survey is complete and does not overlap with any other, and it provides the only method for determining precisely the number of establishments not actually covered by any survey. It is the cornerstone of a comprehensive "annual economic survey".

The recommended procedure is based on the use of a tax-based frame and multiple frame techniques. The reasons for preferring this option to a PD-based scheme are:

(a) income tax data cover non-employers;

(b) the use of income tax data to supplement survey data is greatly facilitated;

(c) by the time survey estimates are being prepared the income tax data will be virtually complete for the reference year and will thus ensure complete coverage of the universe;

(d) weights for the tax based units are computed centrally and hence consistently.

"Prespecification" sampling (ref. Item D) can be used to augment the cross-sectional tax samples in particular industries where the latter by itself might be inadequate. Advantage can be taken of PD data, or any other "signals", to supplement the list of Class I units prior to mailing survey questionnaires.

These general procedures have been used for the Census of Construction for several years.

For consistency and confidentiality reasons it is necessary to define the requirements for industrial and geographic detail in a uniform way.

Strategy

J1. The basic statistical unit for annual surveys of economic production will be the 1980 SIC establishment.

J2. The frames for these surveys will be tax-based for small units (ref. Items C13, C15), and will provided directly by the CFDB system.

J3. The frames will be split into larger (Class I) units and smaller (Class II-IV) units. The former will be subject to a census, the latter to sampling by direct survey and of tax returns.

J4. Survey operations will utilize the CFDB for access and update of universe frame data. They may use CFDB and/or survey specific systems for the sample mailing list.

J5. The survey frames will not be complete at the time of survey mail out. PD data will be used to make them as much complete as possible with respect to Class I units. The frames will provide complete coverage at the time of survey estimation.

J6. The surveys will all collect the same basic data items, including certain frame data (see H2), in a consistent fashion. They will in addition collect many industry specific data from the larger units. Industrial and geographical detail should be specified in a coherent fashion for all annual survey publications although for internal Bureau usage differing levels of detail may acceptable.

J7. Tax data furnished by the CSF from the tax data base will be used to supplement survey data for smaller units.

J8. Estimation will be based on multiple frame techniques [6-10].

Implementation

Implementation of this policy should be on a survey by survey basis. It should certainly take place in the event of a planned major revision or redesign. In total over all surveys a considerable amount of work is involved, too much to be undertaken all at once.

It will be vital to minimize the artefactual impact of CSF and survey changes upon estimates. In particular, support for existing PD-based survey frames should be maintained until the survey changes to a tax-based frame have been fully planned, resourced and implemented (ref. C13, C15). All statistical units which are significant in survey terms should be defined, in so far as possible, as Class I/II members and hence fully linked. Such units will be unaffected by transition to a tax-based frame.

In future developments survey in-sample master files should be implemented within the framework of the CFDB system.

The boundaries below which tax data are utilized to supplement survey data should be determined by comparing the loss of quality (if any) associated with the use of tax data with the reduction (if any) in cost and in response burden.

Some survey-specific implementation notes are given in the following paragraphs.

Census of Manufacturers and Forestry

This survey is currently PD-based, and in a census of all employers with no provision for sampling. Non-employers are mostly but not entirely defined as out-of-scope. There is undercoverage of births for which no PD account has yet been classified and of units which have misclassified to another industry. There are some conceptual differences between this survey and others in the definition of own-account non-manufacturing activity. Adoption of the proposed strategy will help address some of these problems. It may require some fundamental changes.

Census of Construction

This survey is already conducted along the lines of the proposed strategy. Adoption of the strategy will simplify and enhance survey operations. The changeover impact upon estimates and costs will be relatively small.

Transportation Surveys

The annual surface and marine transportation surveys are essentially, PD-based. Establishments below a certain size boundary are ignored and tax data are used only to supplement the large establishment mailing lists. Implementation of the proposed strategy would enhance estimates by providing coverage of smaller units. It would require some additional resources but no really fundamental changes.

Merchandising and Services Surveys

Annual wholesale, retail and services surveys are all presently at various stages in the design/redesign process. All designs are based on the use of tax data. The proposed strategy is one of the design options. Implementation costs would probably be smaller than for alternative strategies in view of the data and systems which would be provided by the CSF.

6. IMPLEMENTATION PLAN:

The elements of the implementation of the strategy have been touched on in the discussion of the various aspects of the strategy. This chapter combines all these elements and discusses (most of) the problems that will be encountered.

A broad outline of the implementation is as follows:

(a) Review (Economic Statistics Programme, this document, BR policy).

(b) Organization (to accomplish implementation).

(c) Review of Economic surveys (frame requirements, impact, change).

(d) Plan the organization of the CSF and plan the transition.

(e) Initialize CFDB (structures, ID, linkages).

(f) Plan and develop maintenance systems for CFDB (linkages, updates, NBE's, survey interfaces, etc.).

(g) Plan and develop procedures for acquisition and processing of Tax Data.

(h) Research on the utility of other files.

(i) Set up contact with RCT.

(j) Linkage of CFDB to other frames used in Economic Statistics.

These will now be examined in more detail.

A. Review

A1. An "Overall Strategy for Economic Statistics" should be developed. This would involve a review of

- user requirements
- data sources and data collection systems
- the current programme and its deficiencies.

The strategy would clearly delineate the global objectives of the Economic Statistics Programme regarding

- reference populations
- concepts and definitions
- industrial and geographic detail required
- statistics to be produced
- frequency of production of the various components
- quality (tolerable error, bias, coverage, etc.)

One can anticipate some of these objectives - the basic data outputs correspond to the current outputs of the principal economic surveys and programmes, but there may be additional requirements due to the introduction of new elements (such as Small Area or Small Business programmes) consolidations, changes in frequency, etc.

The development of such an Overall Strategy would take several years (indeed, it would have to be under continuous review). It is more realistic to address first only those components which will require the CFDB or which will use tax data directly.

A2. A group consisting of personnel representing all divisions in Economic Statistics should review this document (Strategy for Provision of Frame Data and Use of Income Tax Data - "Infrastructure Strategy") and its implications should be thoroughly discussed with the authors. The document would then be modified and disseminated as part of the "Overall Strategy".

This "Infrastructure Strategy" is, in fact, quite limited in scope and should ultimately be extended to include all economic surveys (e.g. farming, governments) and future additions to the programme (see J1).

One may also anticipate that in the review process, the desire to satisfy everyone could lead to the most important benefits being stripped away. A strong mediation process will be required to determine what suggestions are acceptable and which not in the best interests of the Overall Strategy.

A3. The existing BR policy document should be reviewed and replaced by a comprehensive policy identifying the roles of the C.S.F. and its relationship to user surveys. The revised BR policy would incorporate the conclusions reached in the "Overall Strategy" and in the "Infrastructure Strategy".

B. Organization

B1. A steering committee should be set up to monitor the implementation.

B2. A project manager should be appointed.

B3. A working group should assist the project manager to plan the implementation project, i.e. to plan the transition to the goals set by the revised BR policy document and to set up the organization required to do this.

B4. An interdivisional working group should be set up to resolve interdivisional problems, provide subject matter input and facilitate the implementation of the strategy.

C. Review of Economic Surveys

C1. All surveys should be reviewed to determine the frame

requirements in terms of the overall strategy. The appropriate units should be identified and units other than 1980 SIC establishments should be justified.

C2. Surveys likely to be immediately affected by the new "Infrastructure Strategy" should be examined to determine the impact of that strategy regarding:

- population
- concepts and definitions
- stratification
- questionnaire
- methodology (sampling, editing, imputation, estimation, etc.)
- data collection
- data processing
- detail published
- publication
- resources required to run.

It is necessary to identify what these changes will be and what effect the changes are likely to have on the estimates (see (5)). Decisions will have to be made at this time about how tax data will be used and the boundaries below which it will be used to supplement survey data. The possible loss of quality associated with the use of tax data should be compared with the potential reductions in cost and response burden.

In addition, the effects on the continuity of the series must be considered and methods devised to bridge the gap, e.g. by simulating the current survey while running the revision, since a true parallel run may not be possible.

It seems advisable to start this study with a single survey. Subject matter specialists, methodology and CSF or Strategy implementation personnel would work together to produce a comprehensive document. Having produced one Strategy Impact document, it will be easier to produce others and everyone will have a better idea of what is required.

C3. Detailed plans should be drawn up covering the transition of each survey from its current form to the target. These plans will have to be dovetailed into the overall implementation plan so that the infrastructure required by any survey is ready when it is required. Furthermore, it will not be possible to convert all surveys at the same time. The transition will have to take several years, starting with those surveys that are least dependent on the infrastructure. However, any major revision or redesign should incorporate the requirements of the strategy. The possibility of a phased transition will have to be

considered.

C4. In-sample master files should be implemented within the framework of the CFDB system.

C5. Unnecessary change, resulting only from a change in procedures rather than concepts, coverage or data sources, should be minimized. This may be achieved to a large degree by defining (nearly) all "significant" units as Class I/II units, which would leave them unaffected by the transition to a tax-based frame.

C6. Internal and external publicity or information programmes should be developed. No matter how great the improvement in the product, the users need a considerable lead time to prepare themselves for changes, and they have to be convinced that the changes are worthwhile.

It is probably not too much to suggest that the transition will be more smoothly accomplished if <u>everyone</u> involved is persuaded that the effort will result in a better product and that the objectives of the bureau will be better served.

D. <u>Plan the Organization of the CSF and the Transition</u>

Dl. Plan the structure of the CSF, if not in detail, then

sufficiently to know what the basic structure will be and to plan a suitable transition. As the CSF systems are developed, the details of the required final structure will likely become clearer; but the CSF may include units for operations, systems, analysis, planning, quality assurance and administration.

D2. Transition to the CSF will probably require a reorganization affecting several divisions. Central services are currently provided by four separate groups: Tax Record Access, Business Register, Business Finance Corporate Data Base and Business Finance SIC Coding and Rebasing Unit. It appears logical to consolidate these components within the CSF.

This transition should be carefully planned. All personnel affected should be briefed and prepared well in advance so as to maintain good morale.

D3. Transition should begin with the standardization of concepts and procedures, the co-ordination of existing operations and the dissociation of future central services from existing survey specific functions in Business Finance Division.

D4. Transition should be phased. It should be transparent

to user surveys or specifically include the necessary changes as part of the transition plans.

E. Initialize the CFDB

E1. Definitions and concepts relevant to the CFDB must be examined and established, in particular:

(i) definitions of birth, death, transformation;

(ii) the possible "states" of a statistical or administrative unit;

(iii) the relationships between the various sets of statistical and administrative units;

(iv) the system of unit identifiers.

A comprehensive dictionary of terminology should be established to form the basis for the design of storage, updating and integrity-checking procedures.

E2. All source files which are used currently or which are potentially useful for the provision of frame data or tax data should be documented and analyzed systematically to establish the quality of the data and identify potential uses and problems of usage. The analysis should be co-ordinated, performed to a uniform standard and completely documented. It should cover

(i) universe snapshot tabulations.

(ii) assessment of quality of the data on each file (missing data, coverage, data errors).

(iii) longitudinal and between file comparisons, and identification and resolution of differences, where possible.

(See also H1.)

E3. There should be a comprehensive and systematic review of the linkages between units on the administrative and statistical unit files (PD-SIN, PD-T2, T2-ICO, T2-BRID, etc.). A fair amount of linkage information exists, even if the links have not been used. The quality of the linkage needs to be assessed.

The existing links should be extended by a programme of automatic matching, and manual matching for the largest units. All the data of potential value (e.g. size, SIC, "old" linkage information, etc.) should be exploited. On the T1 file, it is necessary to identify all partners in the same business, for the largest businesses. Unlinked large units should be carefully investigated and accounted for. It may be possible to characterize non matches.

The establishment and maintenance of linkages for large and complex units is at the core of the "Infrastructure Strategy". It is thus imperative to demonstrate that

(i) a high percentage of the Class I units can be linked;

(ii) the quality of this linkage is high;

(iii) the linkage can be maintained over time; and

(iv) the cost is reasonable.

Furthermore, the nature of the unlinked units must be thoroughly understood, procedures developed for dealing with them, and the impact of these units assessed.

It is clearly feasible to establish quite extensive linkage. Whether it is feasible to establish a sufficiently high level of linkage in Class I and maintain it remains to be demonstrated.

E4. Information developed in steps E2 and E3 should be sufficient to enable the definitions of Classes I-V to be refined and finalized.

The definition of Class I is most important and has to be delineated in such a way as to be consistent bet-

ween three files.

An appropriate size criterion for the definition of Class I PD accounts must be selected which coincides as closely as possible with the corresponding tax return criteria. It could comprise three distinct threshold boundaries used in hierarchical fashion:

(i) a boundary only in terms of data items appearing on the PD universe file;

(ii) a boundary in terms of PD universe and/or PD20 data items;

(iii) a boundary in terms of PD/T4 data items.

An appropriate size criterion for the definition of Class I tax returns must be selected which corresponds as much as possible with the corresponding PD account criterion, based on studies of matched records. This size criterion might be based on GBI or Wages and Salaries; but should not depend on SIC because of the volatility of the code and because no SIC is on the tax return. On the other hand, the criterion may differ between unincorporated buinesses, professionals and commission agents - this question needs to be examined.

E5. The necessity and feasibility of supplementing the

CFDB with an area frame should be examined. The area frame would serve current surveys for which the CFDB does not provide frame updates on a sufficiently timely basis. On the other hand, it would add considerable additional complexity both to the CFDB and to the redesign of these surveys.

E6. A CFDB/User Survey interface protocol must be established for each user survey. It will specify

(i) the nature and timing of the lists to be provided to the survey;

(ii) the information required by the CFDB from the survey to acquire tax data;

(iii) the nature and timing of the tax data which will be passed to the survey;

(iv) other services to be provided by the CSF (monitoring reports, auxiliary data, etc.);

(v) the procedures to be followed when the CFDB-supplied units turn out to be dead or out-of-scope;

(vi) circumstances which would cause a nature of business inquiry to be initiated;

(vii) data required by CFDB from the survey.

It should also specify arrangements to cover the transition from the current BR/TRA-survey interface to the target interface, as well as the timing of these steps.

E7. Procedures and systems for checking the internal consistency of the CFDB must be developed and implemented.

E8. Sample monitoring facilities must be developed and implemented. Sampling facilities, possibly incorporating response burden constraints, should also be developed in the longer term. The feasibility and effectiveness of such a system needs to be investigated, given the volatility of the business universe, the diversity of survey objectives, the differences in reporting units, etc.

E9. Procedures for the storage, and maintenance of and access to relevant administrative and commercial lists of units and associated data must be developed.

E10. The CFDB must be designed and implemented. The basic components of this data base already exist. However, there are some differences with the current BR, notably

(i) expanded coverage of the file to include both

employer and non-employer units;

(ii) introduction of new stratification variables such as GBI, Assets, Wages and Salaries;

(iii) differentiation between statistical establishment units and legal entities, as different statistical files;

(iv) separation of non-statistical units;

(v) replacement of the hierarchical structure (relationship between units) with a network structure to permit the storage of multiple relationships, and the adoption of an effective new identifier to reflect the new structure.

E11. Procedures for partitioning the CFDB, especially the Class I units, into subframes for the various economic surveys must be developed, using information available from step E6. Surveys which concern themselves with mutually exclusive economic sectors should have mutually exclusive frames, unless there is good reason for them to survey the same units. Surveys which deal with more than one area (such as SEPH) should classify the units in the same way as surveys which deal with specific areas, i.e. the function of co-ordination must be observed.

The task of initializing the CFDB will be very complex and cannot be underestimated. The greatest problem will be to achieve a high quality linkage between the Class I units of the various files and to account for the unmatched Class I units. To date we have experience with the T2-BRID match, but no recent experience with a T1-BRID match or with matching partners in T1 records. It would appear that work in that area could start immediately to establish the feasibility of this operation. The T2-ICO match has not yet been attempted either, but is not quite as crucial.

In order to test the feasibility of initialization and to understand the problems better, the construction of a test subset CFDB should be considered; for example, for a single province, such as BC or Quebec, or the Maritimes, which would allow a reasonable cross section of industries but would suffer from the problem of many multis with only part of the operation in the test province. This subset CFDB would also be of use in developing the various maintenance systems discussed in the next section.

F. Plan and Develop Maintenance Systems for the CFDB

F1. Procedures for the maintenance of established links must be developed. We do not at present have any "hard" information about the stability of the various kinds of links that will be involved. Longitudinal studies will be required to examine this question.

Clearly a complete linkage cannot be done every year and knowledge of which links are stable and which must be routinely repeated is vital to the success of the operation.

Such procedures will include the production of reports covering:

(i) the files being linked;

(ii) the number of in-scope units on each file;

(iii) the percentage of links established at each stage of the linkage process;

(iv) the detailed procedures used at each stage;

(v) the quality of the linkages;

(vi) the nature of the unlinked units, in terms of distributions of GBI, SIC, status, etc..

F2. Procedures and systems for the storage, updating and dissemination of data on the CFDB must be developed. This has to be done in such a way that the different "pieces" of the CFDB "talk" to each other and are automatically kept in step.

In particular, units which appear to be out of scope for their designated surveys must be dealt with in a

timely fashion. Existing procedures should be reviewed and a system to monitor this activity should be set up.

F3. Procedures and systems must be developed for the processing of all relevant administrative data files and the extraction of data to build and maintain the lists of statistical units for specified reference periods.

Investigation of procedures for improved use of PD data for births, deaths and structural changes should continue, together with the analysis and reduction of time intervals involved in processing these data. Current procedures should be extended to make use of PAYDAC and/or T4/T4A for size coding, in particular to identify units which have grown into the Class 1 category.

F4. Procedures and systems which cover the interface between the CFDB and each survey must be developed.

F5. Procedures and systems for editing, cross-checking and monitoring selected data items from various surveys must be developed.

F6. Procedures and systems required for

(i) profiling,

(ii) nature of business enquiries, and

(iii) telephone contacts

must be specified and developed. These would cover the circumstances under which these activities are initiated as well as the actual procedures and documentation involved. The development of a self-profiling survey is also included as well as enquiries which take place as part of a quality evaluation. These activities would all be co-ordinated so that only the minimum number of contacts is made.

The self-profiling survey might be developed along the lines of the USBC survey and using the Quarterly Survey of Corporations/Capital Expenditures as a basis.

F7. Procedures and systems for monitoring the quality of the data on the CFDB must be developed. Some work in this direction has already started.

F8. Confidentiality and security problems associated with the various files must be resolved.

F9. Standard procedures for SIC coding, including the assignment of "degree of certainty" indicators should be introduced. The "degree of certainty" indicator would help in a more realistic assessment of the

quality of SIC coding.

An automatic SIC coding system for smaller units is under development. Its quality, usefulness and cost-effectiveness ned to be assessed.

F10. The conversion of existing establishment units to 1980 SIC should be monitored and schedules modified if necessary.

F11. Plans and procedures for theintroduction of 1980 establishment (statistical and reporting) units must be prepared and implemented in co-ordination with the profiling programme (see F6). Similarly, plans and procedures for the introduction of the 1980 company classification system and company statistical unit should be prepared.

F12. Standardized procedures for SGC coding should be introduced. Software for the automated assignment of SGC from postal codes already exists and will require minor modification.

F13. The Company Contact system should be refined from the existing system.

G. Plan and Develop Procedures for the Acquisition and

Processing of Tax Data

G1. The definition of tax returns in scope for economic surveys must be determined. This will be done in the course of establishing the CFDB/User Survey protocols (E6) and will determine such issues as whether professionals should be included or what the lower bound should be. An analysis of the relative contributions of the smaller units to the totals at provincial and 2-3 digit SIC levels will be required. Because of the size and volatility of the universe, the definition should be as simple as possible.

G2. The definition of "Class I" tax returns must be determined (E4). It should coincide as much as possible with the corresponding PD criterion and should also be relatively stable over time. It can be based only on data items available on the T1 or T2 universe tax files and on the distributions of these items. Analysis of the data from linked PD accounts and tax returns will be required.

G3. Current sampling and estimation strategies should be reviewed, and revised or new strategies determined.

Several programmes have to be accommodated:

(i) the annual surveys of production, including

Services;

(ii) the Business Finance annual survey of corporations;

(iii) the Small Business Statistics programme;

(iv) the Small Area Business Statistics programme;

(v) ad-hoc studies of a cross-sectional or longitudinal nature.

With this in mind, it would be of benefit to:

(i) introduce a T2 cross-sectional sample;

(ii) improve the T1 cross-sectional sample to include stratification by province and finer stratification by size, and to include the sampling of returns showing any form of self-employed income.

(iii) consider sampling by small area (economic region or CMA, using postal code or RCT locality code).

As the various surveys are revised to accommodate the new infrastructure, their tax data strategies may change. Thus the acquisition strategy cannot be expected to be stable for several years, if ever. Furthermore, the strategy will also evolve in response

to the accummulated experience of the various programmes and must be flexible enough to do so.

Research is needed into the levels of precision which various strategies, corresponding to different levels of acquisition, would support for the different programmes. There is some urgency for this. Implementation of a sampling strategy for the current tax year requires immediate action.

G4. Current transcription procedures should be reviewed to accommodate all user needs. This includes the incorporation of the current T2 transcription for Business Finance within the framework of the tax data transcription operation. Account must be taken of any new requirements from the Small Area and Small Business programmes.

G5. Data entry and editing procedures should be reviewed and revised. A system for the production of a weighted T2 sample file (equivalent to the current T1 Sample file) should be introduced.

G6. Data transcribed for the Annual Survey of Corporations should be made available through the CFDB to other user surveys (e.g. for imputation purposes). Conversely, data transcribed for other purposes should

be made available to the Annual Survey of Corporations.

G7. A system for computing the sampling variability of estimates based on the tax samples should be designed and implemented. Such a system is essential

(i) for the development and comparison of sampling strategies, and

(ii) for the production of measures of quality of the estimates produced by the various programmes.

G8. Procedures should be developed for imputation on the tax data files when items are missing from the records. The need for large-scale imputation of detailed transcriptions or survey returns should be examined.

Several systems capable of doing some or all of these tasks are available in the Bureau. Some are used in imputation systems specific to particular surveys. It is important to consolidate the imputation operations so that no two applications which use a single file produce different estimates because of different imputation procedures.

Various imputation methodologies should be evaluated in terms of

(i) the error of the estimates;

(ii) distributional properties;

(iii) ease of use and implementation.

Input would be obtained from all Subject Matter Divisions using tax data, as well as Small Area and Small Business programmes about the desirable properties of imputations for their particular segments.

The resulting imputation system will likely incorporate several types of procedures, but at least all users will use the same files.

G9. An analysis of the incidence and impact of T1 partnerships and multiple businesses must be made, together with an assessment of the performance of the strategy proposed for handling them.

G10. The use of the Cumulative Sample file as auxiliary data should be studied. The use of the Weighted Sample file has been recommended as a basis for annual estimates; but there is some feeling that the Weighted Sample is rather sparse at some levels, while the Cumulative Sample contains "more" records, although

many of these are no longer correctly classified or dead. The question is whether and how to use both of them and to modify the sampling and estimation methodologies accordingly.

A study would probably begin by assessing the quality of the data (in the sense of still being valid) on the Cumulative Sample file, particularly in relation to the age of the record.

G11. Tax data concepts and terminology should be documented and contrasted with survey data concepts. A start has already been made on this in connection with the Small Area Business Data programme, but definitive documentation has yet to be produced.

Comparisons of tax and survey data at the unit level should be made to facilitate the development of a strategy for handling the differences. The impact of defining the annual survey reference year to coincide with the tax reference year should be studied.

G12. The mini-computer system currently being developed within Tax Record Access should ultimately incorporate all sampled tax data. However, this system should communicate with the CFDB.

The tax data requires much manipulation and analysis. In addition to serving the annual surveys, it will also be used in the Small Area and Small Business programs. The ability to work with this data independently of the mainframe should lead to considerable savings of time and computer resources.

H. Investigate the Utility of Other Files

H1. A programme of research should be undertaken on the utility of the various machine-readable RCT files (COMSCREEN, CORPAC, GREENBOOK, PD-SIN, etc.) as

(i) source data, perhaps for updates;

(ii) auxiliary data;

(iii) validation data;

(iv) linkage data.

In particular, the CORPAC file may be of use in identifying new corporations which have not yet filed a return.

Although this Infrastructure proposal does not depend on these sources of data, this does not mean that they could not be directly useful or that any modification of the proposed system to include their use should be ruled out. However, no commitment can be made to use these files until a systematic assessment of their usefulness is carried out.

Some of the problems with the RCT files are well known. However, the structure of these files is subject to change and hence they will have to be constantly evaluated (see Section 6 I).

H2. Depending on the results obtained as a result of H1, modifications or additions to the CFDB system to incorporate information from the RCT files. This implies that the CFDB subsystems will be developed in such a way as to be easily modified.

H3. A policy should be developed regarding the use of the RCT machine-readable files.

H4. A programme of research should be undertaken to evaluate other administrative and commercial data as sources of information for the CFDB which may be more timely or cover areas not covered by tax data. Possible sources of information are

(i) Federal and Provincial lists of corporate charters;

(ii) municipal lists of businesses;

(iii) hydro-electricity accounts, telephone accounts, etc.

The difficulty in using such data is invariably the absence of a common identifier with the consequent risk of duplication.

STC has no experience with such files in connection with the B.R. An evaluation would involve testing various modes of utilization: matching and reconciliation, multiple frame, etc. and an assessment of effectiveness vs cost.

I. Set up Contact with RCT

I1. Set up a joint RCT-STC working group on the use of tax data, to exchange information and negotiate changes in procedures.

I2. Current RCT and STC sampling, transcription and data-capture processes should be reviewed (starting with the Adams-Valiquette study documentation) in order to identify joint (STC and RCT) operational changes which could improve the quality or coverage of tax data, or reduce costs. It is obvious that closer collaboration between RCT COMSCREEN and STC T1 acquisition, for example, would be of mutual benefit.

Areas which could usefully be discussed by the joint working group are:

(i) co-ordination of sampling, transcription and data-capture;

(ii) sharing of sampled data;

(iii) RCT data capture of additional items, such as nature of business information, on a universe basis;

(iv) the assumption of much of the basic transcription by RCT, perhaps on a contract basis;

(v) RCT changes to the tax returns, e.g. introduction of fixed format schedules, prescribed lists of industrial activities;

(vi) RCT co-ordination of PD and tax systems and linkage of identifiers;

(vii) introduction of more flexible sampling facilities;

(viii) RCT use of the 1980 SIC (at least to the 2-digit level) for COMSCREEN;

(ix) RCT changes to the PD-20 questionnaire;

(x) frequency of access to PAYDAC machine-readable files;

(xi) co-ordination of RCT and STC follow up on PD20.

J. Linkage of CFDB to Other Frames Used in Economic Statistics

J1. Linkage between the CFDB and survey frames not derived from it should be established and maintained. The quality of these links should be evaluated.

There are two obvious reasons for this exercise: the improvement or validation of the CFDB and the possible improvement of the survey frames themselves. Less obvious reasons are the control of response burden and increasing the conformity to standardized concepts and definitions.

7. IMMEDIATE TASKS

Introductory Remarks

The Plan outlined in Chapter 6 will take several years to implement fully. This chapter defines a set of tasks which can start now and be substantially accomplished within twelve months. The tasks involve documentation of existing procedures and data sources, comparison of alternative options, preliminary designs, construction of prototype systems and procedures, etc. This sort of preparatory and analytical work will be of intrinsic benefit in testing feasibility and enhancing current procedures whether or not the strategy is fully implemented in its present form.

These tasks will involve many people within BR,TRA, Business Finance, BSMD, SDD, Standards Labour and the Business and Trade Divisions. Parts of some tasks are already under way, other parts are scheduled within the various divisional programmes confirmed in the 1984/85 LTOP. Collectively this set of tasks should be viewed the bridge between the existing programme and a future set of LTOP submissions which incorporates all the work required for full implementation of the Infrastructure Strategy.

List of Tasks

1. Establishing Priorities.

2. Project Planning, Organization and Control

3. Concepts, Classification and Coding

4. Profiling and Self-Profiling

5. Acquisition and Processing of Administrative Data

6. Revenue Canada, Taxation - Bureau Interface

7. Central Service Function and Data Base Design

8. Central Frame Data Base Linkage

9. Use of Income Tax Data to Supplement Survey Data

#10. Review and Redesign of Economic Surveys

#11. Project Operations

Task #1. Establishing Priorities

Objectives:

To define and publicize Bureau priorities relevant to the acquisition and use of frame and tax data.

Output:

A set of criteria indicating the relative importance of the requirements for frame and tax data of the various economic surveys.

Notes:

The task should not really be considered as part of the Infrastructure Project, rather as a prerequisite for its planning and implementation. It requires senior management input and decisions.

In the long term, priorities such as these should, in principle, be obtainable from the Operational Plan Framework (Core/Complementary Analysis) for Economic Statistics.

Cross References: Chapter 5 (A1), Chapter 6 (A1).

Task #2. Project Planning, Organization and Control

Objectives:

(a) to define the terms of reference and objectives of an Infrastructure Project;

(b) to clarify and formalize the strategy as presented in this document;

(c) to discuss the strategy with all parties concerned, to respond to criticisms and queries and make revisions as need be, to prepare a final version of the strategy and have it confirmed as Bureau Policy;

(d) to prepare a plan for implementation of the strategy, defining tasks, resource estimates, schedule, project team structure, milestones, etc.;

(e) to organize the project (obtain resources, identify pracas codes, define project and working group members, identify steering committee or equivalent, etc.);

(f) to control the project (by coordinating the working groups, reviewing and summarizing their output, reporting to the steering committe, etc.).

Outputs: an Infrastructure Strategy, an implementation plan, progress reports, summaries of analyses and of methodological and systems specifications arising from the various tasks.

Note: This task provides the planning, organizational and control framework for the complete set of project tasks (#3-11).
The project team should include senior representatives from BR, TRA, Standards Division, Labour Division, BSMD, SSD and each division within the Business Statistics Branch.

Cross References: Chapter 5 (A2, A8), Chapter 6 (A2, A3, B).

Task #3. Concepts, Classification and Coding

Objectives:

(a) to establish and document all relevant concepts, definitions and terminology;

(b) to complete the strategy by defining target populations, resolving conceptual problems and extending the strategy to cover all aspects of economic statistics;

(c) to review and evaluate, from quality and efficiency perspectives, the existing procedures for SIC coding, geocoding and size coding;

(d) to review and summarize the status of conversion to the 1980 SIC;

(e) to evaluate and enhance software for automated and computer assisted SIC coding, for geocoding and for size coding;

(f) to develop standardized procedures and a coordinated strategy for SIC, geographic and size coding throughout the Bureau, to have these procedures accepted as Bureau policy and to prepare a plan for their implementation.

Outputs: dictionary of terms and concepts; supplements to and extensions of the strategy; documentation and evaluation of current SIC, geocoding and size coding systems and procedures; a status report on the 1980 SIC conversion; Bureau strategy, procedures and implementation plans for SIC, geocoding and size coding; improved SIC, geocoding and size coding software.

Cross References Chapter 5 (C, I), Chapter 6 (E1, F9, F10, F12).

Task #4. Profiling and Self Profiling

Objectives:

(a) to document and review current procedures in relation to the proposed strategy, to revise the proposed strategy if need be and have it confirmed as Bureau policy;

(b) to prepare detailed procedures for profiling by interview, to have them confirmed as Bureau policy and to supplement them with an implementation plan, training material and specifications of the computer processing facilities required for the final CFDB and for the transition period;

(c) to formulate the framework for a self-profiling survey and have it confirmed as Bureau policy; to develop a survey design, procedures and an implementation plan, and to specify the processing and storage facilities required for the final CFDB and for the transition period.

Outputs: summaries of current and proposed procedures; a Bureau policy for profiling and self-profiling; a profiling implementation plan, a self-profiling survey design.

Cross References: Chapter 5 (G), Chapter 6 (F6).

Task #5. Acquisition and Processing of Administrative Data

Objectives

(a) to document and summarize the principal sources of administrative data (both the basic administrative forms/questionnaires and the machine-readable files derived from them), to analyze in each case the quality of the data available and the current procedures for their use.

(b) to prepare a detailed design for processing and use of these data in the proposed CFDB and to identify any potentially major problems;

(c) to prepare an implementation plan and specification for short term enchancements.

Output: summaries of the existing data services and processing procedures; detailed proposals for use of the data in the future CFDB, for transition from currant procedure and for short term enchancements; documentation of potential problems.

Notes. The principal data sources are:

(a) PD: PAYDAC, PD2Ø,T4/T4A;

(b) T1: SEIF, COMSCREEN, Historical Summary File, Combined Master File;

(c) T2: CORPAC, ADMIN FILE, Integrated Master File.

It may be sensible to define subtasks covering each group separately.

The detailed plans should include specification of the units and data items to be processed and stored, specification of the Class I - V boundaries and specification of the procedures for handling T1 partnerships and multiple businesses. They should also include recommendations for changes in RCT processing and RCT/Bureau interface procedures which would improve data quality and/or processing efficiency and suggestions for enhanced use of PD remittance information to improve identification of births, deaths and changes in size.

The uses of tax data to supplement survey data is covered in task #9.

Cross References: Chapter 5 (D, E, F),
Chapter 6 (E2, E4, E9, G1, G2, G8, G9, H1).

Task #6. Revenue Canada, Taxation - Bureau Interface

Objectives:

(a) to review current RCT and Bureau sampling, data transcription, SIC coding and data exchange procedures in relation to the proposed strategy and to identify any major potential problems with the strategy;

(b) to identify changes in current RCT procedures and in the RCT - Bureau interface which would be cost-beneficial;

Outputs: summaries of current sampling, transcription and data exchange procedures; major problems with the proposed strategy; beneficial changes in RCT and RTC - Bureau procedures.

Cross References: Chapter 5(D), Chapter 6 (I).

Task #7. Central Service Function and Data Base Design

Objectives:

(a) to review the current BR, TRA and Business Finance systems and procedures which process and provide frame and tax data, to compare them with the proposed procedures and to identify the required changes and any potentially major problems;

(b) to prepare a detailed design for the proposed CSF and a plan for transition from current arrangements, and have them confirmed on Bureau policy;

(c) to prepare detailed design for the CFDB at successively greater levels of detail including the mechanism for transition from the existing systems.

Outputs: statement of major design problems with and recommended changes to the strategy; CSF and CFDB design documentation; a prototype CFDB and a summary of experience gained from it; a data dictionary, a unit identification system and a statement of source and statistical file interrelationships.

Notes: This task covers all feasibility and design aspects of the CSF and CFDB.

Cross Reference: Chapter 5 (B, C), Chapter 6 (A3, B3, D, E1, F).

Task #8. Central Frame Data Base Linkage.

Objectives:

(a) to review the linkage techniques currently being used and the present state of linkage between the important statistical and administrative files in BR, TRA, MNE and Business Finance Systems subdivisions, to access their applicability to CFDB initialization and maintenance, and to identify any potentially major problems;

(b) to test linkage procedures and analyze the results and hence to identify and put in place the most effective combination of software for CFDB linkage purposes;

(c) to define all the procedures for linkage creation and for linkage maintenance required for the future CFDB and for the transition period.

Output: comprehensive documentation and assessment of all existing linkages and linkage procedures; linkage software matched to CFDB needs; detailed procedures for linkage creation and maintenance; the results of systematic testing of such procedures.

Notes: When and only when linkage maintenance procedures for the future CFDB and for use during the transition period have been determined and tested should linkage on a large scale for the CFDB be initiated. This does not preclude experimentation to check feasibility, or to help establish the likely Class I/II/III boundaries, or current routine operations.

Cross References: Chapter 5 (D3,E2,I8), Chapter 6 (E).

Task #9. Use of Income Tax Data to Supplement Survey Data

Objectives:

(a) to review current procedures for the use of tax data to supplement survey data in relation to the proposed strategy, to identify any potentially major problems, to revise the strategy as need be and have it confirmed as Bureau policy;

(b) to prepare detailed plans for sampling, transcription, processing and usage of tax data in the short term, for the transition period and in the long term.

Output: summaries of existing procedures for sampling, transcription, processing and usage of tax data for surveys; problems with and recommended changes to the proposed strategy; detailed proposals for the final system, for the transition period and for short term enchancements.

Cross References: Chapter 5 (D), Chapter 6 (G).

Task #10. Review and Redesign of Economic Surveys

Objectives (for each economic survey):

(a) to review, from a Bureau perspective, the survey objectives and the frame data requirements to which they give rise; to relate item to the proposed strategy for provision of frame data and use of tax data; to identify any potentially major problems in implementation of the strategy and to suggest revisions;

(b) to prepare a detailed plan for transition from current procedures to those proposed;

Outputs: (for each economic survey); documentation of survey objectives and consequent frame data requirements; problems with, revisions to the proposed strategy; an implementation plan.

Note: Preparation of an implementation plan should be primarily aimed at establishing feasibility and revealing problems. Implementation may have to wait several years for some surveys.
For subannual surveys the potential utility and cost of a supplementary area frame should be considered.

Cross References: Chapter 5 (J), Chapter 6 (C).

Task #11. Project Operations

Objectives:

(a) to establish and maintain good communication between the project team, the project working groups, the subject matter divisions and other related projects (especially the small area business statistics development);

(b) to control the project on a day to day basis.

Outputs: progress reports.

Notes: This task is an extension of task #2 requiring the attention of a dedicated working group.

Cross References: Chapter 6 (B).

LIST OF REFERENCES

[1] Standard Industrial Classification 1980. Statistics Canada, Catalogue 12-501E.

[2] Provision of Frame Data for Business Surveys Draft for Discussion (January 1982), Summary (April 1982), M.J. Colledge, BSMD.

[3] Redesign of Business Measurements, T. Gigantes, January 1984.

[4] BR/DMF Evaluation. J.E. Doucet, SDD, August 1980.

[5] Business Register Policy Statement. S. Ostry, 1973.

[6] Estimators Based on Several Stratified Samples with Applications to Multiple Frame Surveys, M. Bakier, BSMD, 1984 (to appear in the Journal of the American Statistical Association).

[7] Census of Construction Tax Year 1980 Initial Weights, M. Bankier, BSMD, February 1982.

[8] Impact of Non-Probability Samples on Variance of Multiple Frame Estimators, M. Bankier, BSMD, March 1984.

[9] Estimating the variance of a Raking Ratio Estimator by Linearizing Sample Data, M. Bankier, BSMD, September 1983,

[10] Unbiased Estimator of a Variance Formula for an Estimator based on any Number of Independent Stratified Samples, M. Bankier, BSMD, Jul

[11] Report 1. The Existing System of Collecting Data for T1 Tax Returns. R. Adams (RCT), M. Valiquette (STC), November 1981.

[12] Report 2. The Existing System of Collecting Data for T2 Tax Returns. M. Valiquette (STC), R. Adams (RCT), December 1981.

[13] TY84 Project (Memo), A.S. Foti, February 1983.

[14] Company Classification System: Classification Structure, Standards Division, 1983.

[15] Generalized Iterative Record Linkage System - Strategy Guide, T Hill, F Pring-Mill, EDP Planning and Support, May 1984.