PHARE 2002
Multi Beneficiary Statistics Programme
(Lot 1[1])

Quality in Statistics

Task T3: Development of Methods for Quality Assessment

Coverage Procedures for Economic Surveys and Methods for Assessment of Their Quality

June 2005

prepared by

Michael Colledge
Project Expert

[1] Pilot Project 1 of the European Community's Phare 2002 Multi Beneficiary Statistics Programme (Lot 1) is devoted to "Quality Assessment of Statistics" in ten Beneficiary Countries, namely Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia.

1. Introduction

The primary aim of the paper is to discuss coverage procedures for statistical surveys and methods for measurement of their quality. The focus is on economic surveys though there is some discussion of household based surveys in so far as they are a source of economic statistics.

The paper contains 11 sections, organised in three parts as follows.

Part A: General Background

To give perspective and to ensure that there is common understanding of the concepts behind coverage issues, the Part A of the paper provides an extensive background.

- Section 2 presents the basic terminology and outlines the various types of survey and associated coverage issues.
- Section 3 describes the conceptual framework that underlies an economic statistics programme.

Part B: List Based Enterprise Surveys and Statistical Business Register

Given that most economic data are collected by list based enterprise surveys and that a good quality statistical business register is the primary mechanism for ensuring their coverage, Part B of the paper focuses on the business register.

- Section 4 describes why list based enterprise surveys form the bulk of an economic statistics programme.
- Section 5 summarises the reasons for statistical business register and describes its construction and maintenance, explaining why a statistical business register should exist and on what principles it should be constructed and maintained. The description is based on international recommendations and regulations, to the extent that they exist, including EC regulation CR 2186/93 and the Eurostat Manual that accompanies it.
- Section 6 discusses the functions of a statistical business register.
- Section 7 describes statistical business register organisation, procedures, and systems.
- Section 8 summarises the performance and quality indicators directly or indirectly associated with coverage of list based enterprise surveys. Collectively these indicators constitute a comprehensive *best practice assessment template* for coverage.

Part C: Other Types of Surveys, Collection and Coverage Issues

- Section 9 briefly describes other types of surveys, specifically including those based on area frames.
- Section 10 outlines coverage procedures for surveys based on area frames.
- Section 11 contains some notes on web based surveys with reference to coverage.

Part A: General Background

2. Statistical Framework and Terminology

2.1 Introductory Remark

The material in this section is presented primarily to promote a common understanding of the terms we are using, in particular what is meant by *coverage*, and a common view of the statistical framework within which statistical data, especially economic data, are collected, processed, and disseminated.

2.2 Economic Statistics Programme

The broad level user requirements for economic data, in particular the data required to compile the national accounts, determine the *populations* and *data elements* that comprise the target data outputs of an economic statistics program.

There are two basic methods of data collection, namely from administrative sources and by survey. In either case the data originate primarily in the accounting records of enterprises. Administrative sources are less costly but are not designed primarily with statistical data requirements in mind. Thus collection of data by survey is inevitable.

The System of National Accounts 1993 (SNA93) provides a good indication of the populations from which data are to be collected. However, there are no international standards defining exactly how data elements should be grouped for survey or dissemination purposes. The SNA93 does not specify any data collection methods. The EC regulations provide some guidance. Typically, data elements are grouped by source and/or subject matter under headings such as production, capital expenditure, employment, stocks and sales, retail trade, foreign trade, prices, money supply, etc. They may also be grouped according to the frequency with which they are required and the reference period or point in time to which they refer.

2.3 Survey and Frame

A *survey* is collection of data for statistical purposes. The term survey includes a *census* in which all units are selected as well as a sample survey in which only a sample of units is selected. A survey may be conducted once, or repeated at regular intervals or at irregular intervals. Repetitions of a survey with essentially the same objectives and methodology are referred to as *survey occasions*, and the whole set of survey occasions is deemed to constitute the survey.

The objectives of a survey are expressed in terms of a set of *data elements* (or *data items* or *indicators*) associated with a *target population* (or *target set of observation units).* Building on the international metadata standard ISO 11179, these concepts may be more precisely defined as follows.

A *data element* is defined as in terms of three component elements: *object class, property* and *representation.*

A *property* is "a peculiarity common to all members of an object class". The notion of property is separated from its *representation*. If an *attribute* of an object class is defined as a mapping from the object class into a value domain, then a property is an attribute without specifying the value domain.

- For example, sales is a property regardless of whether the corresponding values are expressed in Australian $, '000s Australian $, US $, or a range such as small/medium/large.

There are hundreds of thousands of properties. The challenge is to define and identify them efficiently.

A population is defined in terms of an *object class* and a set of *inclusion criteria*.

ISO 11179 defines an *object class* as "a set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behaviour follow the same rules". There are typically a relatively small number (meaning dozens rather than thousands) of object classes relevant to economic surveys.

- Particular examples of object classes are enterprise, enterprise group, type of activity unit, establishment, building approval, motor vehicle registration, export transaction, etc. As discussed later, there is a degree of subjectivity in their choice.

The goal in determining the set of object classes to be used by surveys is to minimise their number whilst preserving their homogeneity and ensuring that populations drawn from them can be enumerated.

A *population* for a *survey* is defined as the set of objects (more commonly called *units*) belonging to a particular *object class* in accordance with specified *inclusion criteria*. The criteria must include *geographical* and *temporal* criteria and may include others.

- An example is of a population is {enterprises, in Australia, economically active, any time during fiscal year 2003/04, whose primary activity is manufacturing} where *enterprise* is the object class and the other items are the inclusion criteria.

It may be impossible to assemble the complete target population, and this leads to the concept of a *survey frame* (or *sampled population*), comprising the *sampling units* from which the survey sample is actually selected, and the *reporting units* from whom the data about the target observation units are actually collected. The difference between the target and sampled population is usually reflected simply in the inclusion criteria. Occasionally, the object classes are different and in this case a transformation between the object classes must also be defined.

The survey frame is one of the starting points in conducting a survey. It includes not only the units subject to sampling but also the attributes of those units, known as *frame data items*, that are required in order to identify the relevant population, to stratify and select the sample and to contact the sample units to collect data. These frame data items include:

- identification and contact data – identification code, name of unit, addresses, contact persons
- classification data – legal form, institutional sector, economic activity, size, region;
- demographic data – date of creation, incorporation as a legal person (if applicable), change of structure or activity; activity status indicating whether the unit is active, dormant, defunct, etc;
- linkage data – links to related units (for example, local units to the enterprise, enterprises to the legal units that own them), links to data about the same unit in other files.

The survey frame has more influence than any other aspect of survey design upon the coverage of a survey.

2.4 Types of Surveys

Economic surveys may be divided into five types according to the units (objects) sampled and/or contacted, namely *enterprise surveys*, *household surveys*, *mixed household-enterprise surveys*, *indirect enterprise surveys*, and *price surveys*.

Enterprise surveys are those in which enterprises (or statistical units belonging to these enterprises) constitute the sampled units, the reporting units and the observation units. By

contrast, in *household surveys* the sampled, reporting and observation units are households. In *mixed household-enterprise surveys* the sampled units and initial reporting units are households but the final observation units are enterprises. In *indirect enterprise surveys*, the reporting enterprises are asked for data about a different set of enterprises, i.e., the observation units do not belong to the reporting units. An example would be a survey of city markets in which the market administrators are asked about the numbers and turnover of the market traders. *Price surveys* are those used to obtain data on prices, which may involve collection from enterprises or households, or direct observation of prices in the market.

Surveys may also be classified as *list based* or *area based* depending upon the source of the list of enterprises or households from which the survey sample is drawn. In a list based survey, the initial sample is selected from a pre-existing list of enterprises or households. In an area based survey, the initial sampling units are a set of geographical areas. After one or more stages of selection, a sample of areas is identified within which enterprises or households are directly listed. From this list, the sample is drawn and data obtained.

In most countries with well developed statistical systems, list frame based enterprise surveys constitute the bulk of the economic statistics program. Thus, they are the focus of this paper.

2.5 Scope and Coverage

Coverage is defined in the United Nations Glossary of Classification Terms as

- *the population from which observations for a particular topic can be drawn. An understanding of coverage is required to facilitate the comparison of data. The rules and conventions of coverage are largely determined by concept definitions, scope rules, information requirements and, in the case of statistical collections and classifications, collection and counting units and the collection methodology.*

More precisely the *scope* of a survey refers to the *target population*, the *coverage* of a survey to the *sampled population*, i.e., the *frame*.

For economic surveys conducted within the framework of the SNA93, coverage relates to production activities within the production boundary but is translated, for survey purposes, into coverage of the units undertaking these activities.

Non-response is not considered an aspect of coverage. It is survey characteristic in its own right.

A single, summary measure of coverage is the proportion of units belonging to the target population actually covered by the frame. More detailed measures of coverage include that:

- the frame for a survey should contain all the units that are in the survey target population, without duplication or superfluous units;
- associated with each unit should be all the data items required for efficient stratification and sample selection; and
- the contact information, and this data should all be accurate and up to date.

These notions are elaborated in a later section.

3. Conceptual Framework for Economic Statistics

3.1 Need for a Conceptual Framework

Economic data are acquired through a suite of economic surveys and collections from administrative sources. Typically the data are collected, processed and analysed by separate organisational units, each focussing on a particular subject matter area. Harmonisation and integration of data from the various areas depends upon the definition and use of a common

conceptual framework. Thus, there is a conceptual framework underpinning every integrated economic statistics programme. *Understanding this framework is a prerequisite for a comprehensive description and assessment of coverage.*

The conceptual framework should be based on international standards wherever available to ensure international comparability of the resulting statistics. The System of National Accounts 1993 (SNA93) (or its compatible European extension ESA95) is the starting point and most important harmonising and integrating mechanism. It defines the *economic production* activities in which *institutional units* in their role as *enterprises* engage.

3.2 Economic Production, Institutional Units and Enterprises

The SNA93 actually defines two concepts of economic production: the broad concept of general economic production, and the narrower concept of economic production for SNA93 purposes. *General economic production* is an "activity carried out under the control and responsibility of an institutional unit that uses inputs of labour, capital and goods and services to produce outputs of goods and services". "All goods and service produced as outputs must be such that they can be sold on markets or at least be capable of being provided by one unit to another". Thus purely natural processes such unmanaged growth of fish stocks and basic human activities such as eating and sleeping are excluded. *Economic production for SNA93 purposes* is the subset of general economic production obtained by excluding all household activities that produce domestic or personal services for final consumption within the same household except housing services provided by owner-occupiers to themselves and services produced by paid domestic staff.

An *institutional unit* is defined as "an economic unit that is capable, in its own right, of owning assets, incurring liabilities and engaging in economic activities and in transaction with other entities" (SNA93: 4.2). "There are two main types of units in the real world that may qualify as institutional units, namely persons or groups of persons in the form of households; and legal or social entities whose existence is recognized by law or society independently of the persons, or other entities that may own, or control them." (SNA93: 4.3)

An *enterprise* is an institutional unit, i.e., corporation, government, non-profit institution or household, in its capacity as a producer of goods and services (SNA93: 5.1).

- A *corporation* is "a legal entity created for the purpose of producing goods or services for the market, that may be a source of profit or gain to its owner(s); it is collectively owned by the shareholders who have the authority to appoint directors responsible for its general management". The essential features of a corporation are its creation by law, its accountability for its own actions, its ownership and control by shareholders, and its purpose to produce goods and services at economically significant prices (SNA93: 4.23-24).
- A *quasi-corporation* is an unincorporated enterprise (defined below) that operates as if it were a corporation and thus must be treated as if it is one. Typically this includes enterprises owned by non-resident or government institutional units (SNA93 4.49-53).
- *Non-profit institutions* are "legal or social entities created for the purpose of producing goods and services whose status does not permit them to be a source of income, profit or other financial gain for the units that establish control or finance them." (SNA93 4.54)
- All other enterprises are *unincorporated enterprises*. They may be conducted by governments or households. Household unincorporated enterprises may be considered in two groups: *household unincorporated market enterprises* producing goods or services for sale or barter and *household unincorporated enterprises producing* (primarily) *for their own final use* (SNA93: 4.144, 147). Many households do not contain an unincorporated enterprise. Some may share an unincorporated enterprise in a partnership.

In summary, the SNA93 definition of enterprise is very broad. By including households producing goods entirely for their own consumption, it goes far beyond what many economic survey statisticians would regard as an enterprise. However, the definition is justified if a single term is to be used for all the units engaged in productive activities. An alternative and equivalent term in common use is *economic entity*. This latter term is preferred in subsequent chapters of this report as *enterprise* in the NSC tends to be reserved for legal person engaged in market operations. The term *business* is also sometimes used as a synonym for *enterprise* or, more frequently, for *market enterprise*.

3.3 Enterprise Registration Requirements

The legislation and regulations under which enterprises are legally obliged to register or report are very important from the practical viewpoint of maintaining lists of enterprises for statistical purposes. Examples of such regulations are income tax acts, social security acts, factory acts, and employment payroll deduction acts. They differ considerably from one country to another. Their application can vary according to the legal status of an enterprise, its activities, and sometimes its size.

Each regulation results in an *administrative business register* of enterprises or related units bound by that regulation and in transaction data resulting from application of the regulation. Usually the register and data are referred to collectively as an *administrative source*. Each administrative business register is potentially useful in creating and maintaining the statistical business register, as described shortly.

3.4 Classification of Enterprises for Data Collection and Analytical Purposes

The inputs, activities and outputs of enterprises are heterogeneous. Thus, for analytical purposes, a breakdown by economic activity is invariably required, and breakdown by region and/or institutional sector is useful. In addition, for sampling and data collection purposes, and for some analytical purposes, breakdown by size is required. The classifications that provide these breakdowns are described in the following paragraphs. For each classification, the categories must be mutually exclusive and exhaustive to ensure complete coverage of the total economy without duplication.

Economic Activity. To enable cross-country comparisons, it is essential to use an international standard for classification by economic activity. The International Standard Industrial Classification of All Economic Activities, Revision 3 (abbreviated ISIC Rev 3) is the internationally recommended standard (SNA93: 5.3, 5.5). It is used to classify each enterprise (or part of the enterprise, as further elaborated below) according to its primary activity. The *primary activity* is defined in that activity that generates the most value added (SNA93: 5.7). NACE Rev 1 is the compatible EU equivalent.

Geography. There is no single international standard for geographical classification as countries are quite different in shape and size. The European Council nomenclature for territorial units provides guidance. The most important principle is that, within a country, statistical measurements should all be based on a single national classification. The factors to be taken into account in selecting or designing this national standard are:

- user needs for geographic breakdown;
- the area boundaries that are of most use for sample stratification and data collection; and
- existing administrative boundaries – in addition to the fact that users may require data for administrative areas, it is cheaper for the NSO if another organisation is responsible defining and maintaining the geographical descriptions.

Size. In comparison with people, and even households, enterprises are very heterogeneous. In particular, they can vary enormously in size. Thus classification by size is vital for data collection purposes, and useful for analysis. The EC Regulation 58/97 on structural business

statistics specifies a classification by size, but there is no universal standard. Current practice suggests that, for most data collection purposes, a classification into four basic size groups based on number of employees is useful:

- Micro: from 0 to x employees – x may be 0 or as large as 5
- Small: between x+1 and y employees – typically y is in the range 20-50;
- Medium: between y+1 and z employees - typically z is in the range 50-500;
- Large: more than z employees.

Sometimes no distinction is made between small and micro and/or between medium and large. The boundaries may vary according to the economic activity group of the units being classified, for example to ensure that a given percentage of activity within the group is covered by the large and medium size categories.

Institutional Sector. A classification of institutional sector is required for national accounts purposes. It should be coincide with the SNA93 classification of institution units.

Type of Ownership/ Legal Form. There is no international standard. A national standard should be defined and used.

Activity Status. There is no international standard. A national standard should be defined and used.

3.5 Statistical Units Model

The SNA93 and ISIC Rev 3 recognise that an enterprise, particularly a large enterprise, may be engaged in a range of different activities at a set of different locations. In such a case, the classification of a large enterprise to a single activity at a single location results in a blurring of detail that would be useful for analysis purposes. This leads to the concept that large enterprises should be divided into smaller, more homogeneous *producing units* that can be more precisely classified and that collectively represent the enterprise as a whole.

According to the SNA93, partitioning an enterprise by reference to its various economic activities results in one or more *kind of activity units* that collectively represented the enterprise as a whole. Partitioning an enterprise by reference to its various locations results in one or more *local units*. Using both methods of partitioning simultaneously results in one or more *establishments*. More specifically, an establishment is defined as "an enterprise or part of an enterprise at a single location, engaged in essentially a single activity, and capable, in principle, of providing the data required for the production and generation of income accounts" (SNA93: 5.21-27). The notion of partitioning may be further extended to produce subdivisions of an enterprise into other units each capable of supplying various different groups of data items. For example, there may be a breakdown into somewhat larger units than establishments for which balance sheet and financing data are available, or, somewhat smaller units for which employment and earnings data can be obtained.

The definitions of a set of *standard statistical units* thus created together with the underlying institutional and legal units is commonly referred to as the *statistical units model*. Despite the widespread agreement on the use of SNA93 and ISIC Rev 3/NACE Rev 1 and their European Commission equivalents, there is a range of different units models in current use across the world. The European Council Regulation on Statistical Units 696/93 defines a legal unit and eight statistical units, namely enterprise, institutional unit, enterprise group, kind of activity unit, unit of homogeneous production, local unit, local kind of activity units and local unit of homogeneous production. A *legal unit* is an entity that has *legal personality*, meaning the right to ownership, to dispose of assets, to engage in activities and to enter into contracts and to institute legal proceedings. A legal unit may be either a *legal person* or a *natural person*.

The key principle in the choice of standard statistical units model is that it should be as simple as possible whilst providing detail sufficient to meet user needs. There are two reasons to aim

for simplicity. The first is the high cost of dividing enterprises into smaller producing units. The second is the high cost of building and maintaining a database that is capable of supporting a complex statistical units model. The simplest model of all is to define no breakdown of enterprises at all and, for better partitioning of the data, to ask each enterprise to report its principal data items broken down by kind of activity and location. However, in effect, this is asking each enterprise to make its own breakdown without providing guidance how to identify the units into which the enterprise's activities should be subdivided. It does not work well as an enterprise cannot be expected to understand the breakdown that is required.

3.6 Non-Observed Economy and Informal Sector

Activities that are missing from the basic data used to compile the national accounts because they are underground, illegal, informal, household production for own final use, or due to deficiencies in the basic data collection system are referred to as *non-observed*. They are said to comprise the *non-observed economy (NOE)*, and including them in the national accounts is referred to as *measurement of the NOE*. Undercoverage, non-response and underreporting are the major contributing factors to the NOE. Coverage of the informal sector poses particularly difficult problems.

An explicit designation of certain enterprises as informal is not required for national accounts compilation purposes. However, the informal sector represents an important part of the economy in many countries, playing a major role in production, income generation and employment creation. Thus, its definition and measurement are important. The SNA93 recommended the use of Resolution II of the Fifteenth International Conference of Labour Statisticians 1993 (SNA93: 4.159). The Resolution (abbreviated ICLS93) describes the informal sector in broad terms, and provides the framework within which each country must formulate its own specific operational definition.

"The informal sector may be broadly characterised as consisting of units engaged in the production of goods or services with the primary objective of generating employment and incomes to the persons concerned. These units typically operate at a low level of organisation, with little or no division between labour and capital as factors of production and on a small scale. Labour relations - where they exist - are based mostly on casual employment, kinship or personal and social relations rather than contractual arrangements with formal guarantees." (ICLS93: 5(1)).

More specifically, the informal sector is defined as a subset of household unincorporated enterprises, comprising those enterprises that:

- produce at least some output for the market; and
- are less than a specified size in terms of the number of persons engaged, or of employees or of employees employed on a continuous basis; and/or
- are not registered under specific forms of national legislation, such as factories or commercial acts, tax or social security laws, professional groups' regulatory acts, or similar acts, laws or regulations established by national legislative bodies.

In addition enterprises that are involved in agricultural production may optionally be included or excluded.

These criteria do not result in exactly the same definition of the informal sector across countries. For example, the criteria can be applied in different combinations, the national legislations may differ, or the employment size limit or how it is measured may vary. The Delhi Group has been trying to narrow down the options and thus improve international comparability.

Part B: List Based Enterprise Surveys and Business Register

4. Use of List Based Enterprise Surveys

In a list based enterprise survey, the initial sample is selected from a pre-existing list of enterprises. Typically the list is supplied from the business register (described later) that is maintained by the statistical office to support a range of surveys. Sometimes the survey list is derived from another administrative register. In an area based enterprise survey, a sample of areas is selected, within each of which enterprises are enumerated and then sampled. List based enterprise surveys are generally preferred to area based surveys for the following reasons.

- A list based survey is more efficient from a sampling perspective. Because the area based approach involves cluster sampling, a larger sample is required to achieve a given level of precision than in the case of list based survey.
- It may be difficult to enumerate the enterprises within an area. While retail enterprises are likely to be readily visible, service enterprises that carry out their work in other locations are usually difficult to identify.
- Maintenance of a list of enterprises via a general purpose business register is cheaper than maintenance of an area based list, except for very small enterprises;
- Area based sampling is inappropriate for large or medium sized enterprises that operate in several areas because of the difficulty of collecting data from just those parts of the enterprises that lie within the areas actually selected. Furthermore, in order to avoid inadvertently missing parts of the enterprise, it is usually considered preferable to collect data from the whole of an enterprise not just a part of it.

Thus, area based enterprise surveys are typically only used for collection of data from small enterprises (particularly agricultural smallholdings) and then only when no adequate list exists. Even in these circumstances, a mixed household-enterprise survey (described below) may be preferable. Those area based enterprise surveys that do exist are usually supplementary to a list based survey. An example is the long-standing area frame component of US Bureau of the Census Retail Trade Survey.

5. Statistical Business Register Construction and Maintenance

5.1 Need for a Statistical Business Register

Coordination of the individual surveys and administrative collections constituting the economic statistics program is vital, particularly given that in many cases the surveys were established more or less independently of one another at various different times. Coordination depends upon the use of a common conceptual framework of the type just described. A crucial aspect of making this framework operational is to ensure that the individual survey frames are properly harmonised.

An up to date frame is required for each repetition of a regular survey. It is more efficient to maintain and update the survey frame so that it can support the sequence of repetitions than it is to create the frame afresh with each repetition. This is particularly true in the case of monthly and quarterly surveys where overlap of sampled units from period to period is essential. Frame maintenance is best achieved through the construction of a single *statistical business register (SBR)* and its use as the source of frames for all list based enterprise surveys. Where the context is clear the term *statistical business register* is often abbreviated to

statistical register or *business register* but in the latter case statistical and administrative roles must be clearly distinguished as described below.

There are three basic reasons why a SBR is desirable.

- First, and most importantly, if survey frames are independently created and maintained there is no means of guaranteeing that they are co-ordinated. There may be unintentional duplication or omission of enterprises in the coverage of the suite of surveys.
- Second, a SBR enables practical application of a standard framework of units and their classifications. The starting point for thinking about a SBR is the conceptual framework. Conversely, the SBR helps operationalise the conceptual framework.
- Third, it is more efficient for a single organisational unit within the national statistical office to be responsible for frame maintenance than for each subject matter unit to maintain its own frame.

The only surveys that need not be based on a SBR are those where the frame is derived from a well defined regulatory process and for which there is no requirement for co-ordination with other surveys. An example might be a survey of registered banking organisations that collected financial information specific only to the banking industry.

5.2 Administrative Registers

In principle a SBR can be built from scratch and maintained by enumeration of all enterprises in the country. However, this is an impossibly expensive process. Thus, the starting point for a business register is invariably one or more *administrative (business) registers*, i.e., registers of enterprises that are created and maintained to support the administration of regulations, and to which the statistical office has access under its legislation. The ideal administrative register would be one that provided complete, up to date coverage of all enterprises within the SNA93 production boundary, without duplication or inclusion of defunct units, and that provided all the frame data (classification and contact items) required for sampling and data collection. However, given the very broad range of enterprises within the production boundary (including unincorporated household enterprises, even those with no market output) there is no such perfect source in practice.

5.3 Coverage of Statistical Business Register

The choice of the administrative register or registers upon which to base the SBR varies from country to country. In some countries, for example in France, the administrative business register used to underpin the SBR results from a regulation that specifically takes statistical needs into account and that is actually administered by the national statistical office. In such cases the SBR and administrative business register may actually be housed in the same database, share data items and be referred to jointly as the *business register*. SBR coverage and content are likely very good. In other countries use is made of an administrative register maintained for other purposes, for example, value added tax in New Zealand and chamber of commerce enrolments in The Netherlands. The resulting SBR is inherently likely to be somewhat more deficient in coverage and content.

Given that no one administrative register provides complete coverage and content, combining data from several administrative sources is an improvement option. However, use of multiple administrative registers is practical only if they are known to contain mutually exclusive sets of enterprises or if they share a common identification scheme that allows records for the same enterprise to be brought together and thus not duplicated. Experience has shown that trying to match units across registers in the absence of a common identifier is impossibly expensive unless one of the registers is very small. Furthermore, the gain in coverage resulting from the incorporation of data from an additional administrative source may not

justify the increase in cost, particularly if the additional source is of poor quality. Thus use of multiple administrative sources should be undertaken with caution.

EC Regulation CR 2186/93 specifies some groups of enterprises that should be excluded from a SBR on account of their small size, namely household unincorporated enterprises that are:

- producing for own final use, including those producing domestic services within the meaning of NACE Rev 1 Section P (i.e. households as employers of domestic staff)
- owning property and earning rental income from it, i.e., NACE Rev 1 Group 70.2.

The Regulation also states that, optionally, enterprises engaged in NACE Rev 1 Section A (Agriculture, hunting and forestry) and/or Section B (Fishing), and/or Section P (Public administration) may be excluded (except for those that have significant activity in another section). Certainly enterprises in these sections, and in finance, need special treatment as the sources of data for them are quite different than for enterprises in other sections. Hospitals and schools also pose a particular problem as they can be both private and public.

5.4 Profiling Large Enterprises

An administrative register provides a list of legal entities, or some breakdown of these entities, to suit the administrative purpose for which it is designed. Typically it does not break enterprises down into smaller producing units, classified by economic activity, in accordance with the statistical units model. This must be done by business register staff using a process commonly referred to as *profiling*. Profiling procedures include the rules for identifying all the units defined in the statistical units model.

Profiling of large enterprises often requires personal visits by statistical office staff to gather the required information and tends to be expensive and resource intensive. Thus it is typically restricted to large enterprises. For two reasons, it is rarely worthwhile subdividing medium or smaller enterprises, even if they are engaged in a variety of activities. First, the enterprise may not actually be able to report data for subdivisions of itself; and second, the loss of information by not subdividing the enterprise is usually statistically insignificant and/or less than the errors in trying to obtain a subdivision.

5.5 Content of Statistical Business Register

The administrative registers chosen as the basis for the SBR inevitably have other deficiencies apart from coverage. In particular, the frame data items required may not all be available, or they may be wrong or out of date. To the extent that sources allow, these deficiencies must be addressed through *register improvement surveys* (sometimes called *nature of business surveys* or *register proving surveys*) conducted by the business register staff on a continuous basis. As further discussed below, register improvement surveys are specifically aimed at clarifying or supplementing the information obtained from the administrative sources and hence improving business register quality.

5.6 Maintenance - Dealing with Enterprise Dynamics

Enterprises do not remain the same over time. The institutional units that own them may merge or amalgamate; they may split up or go out of business; they may change production activities, they may move location; and so on. New enterprises may be created (*births*); enterprises may cease to exist (*deaths*); and ongoing enterprises may change activity. Births, deaths, and changes of classification of enterprises must all be fully defined, and the corresponding SBR procedures must be articulated. For example, it must be clearly stated whether an enterprise can be deemed to continue existence through a change of ownership, or whether a change of ownership inevitably means the death of the enterprise record and the birth of another. For practical reasons, these procedures depend upon the sources of information for updating the SBR, of which there are three basic types:

- the administrative sources on which the SBR is based;
- feedback from enterprise surveys carried out using survey frames extracted from the SBR; and
- register improvement surveys directly conducted by SBR staff specifically to maintain the SBR.

Given the large number of small enterprises in any market economy, it is vital that maintenance of the SBR is automated to the maximum extent possible. This means that the frame data for small enterprises should be maintained for the most part by updating the SBR from administrative sources. Updating must be substantially automated as there are neither the time nor resources for SBR staff to verify all frame data received from each source. Staff effort should be focussed on collecting and verifying frame data for the medium and larger enterprises.

Administrative registers are notorious for containing inactive units. Thus, it is vital to make use of any information from administrative sources that can indicate whether the enterprise is active or not. For example, if an administrative source contains information about enterprises required to make payroll deductions on behalf of employees, then the date of the last recorded deduction and the total size of the deductions over the preceding year are good indicators of enterprise activity. No deductions at all suggest that the enterprise is inactive, at least as an employer. This information can used to reduce the number of inactive enterprises wrongly classified as active in the SBR. For example, at the Australian Bureau of Statistics if there is an absence of any transactions recorded by the Taxation Authority business activity statement for five quarters the corresponding enterprise is classified as inactive.

Administrative data will not contain all the classification and contact information required to keep the SBR up to date. For example, it may be impossible to track an unincorporated enterprise through a change of owners as the sale of the enterprise may well appear in the underlying administrative register as the death of an enterprise and the birth of another.

Feedback of frame data obtained from the enterprise surveys is another vital updating tool. There are no technical problems with using data concerning enterprises that have been sampled with certainty - typically the large and medium size ones. However, for small enterprises that are sampled with probability less than one in repeating surveys, the updating information has to be carefully applied so as not to cause bias in future survey samples. For example, suppose that when a particular quarterly survey is first conducted, the sample is found to contain 30% dead enterprises. (This is not an improbable figure.) Furthermore, suppose that, based on this sample information, the dead enterprises are removed from the SBR, and that the survey sample for the next quarter comprises the 70% live units from the previous sample plus a replacement of the 30% drawn afresh from the SBR. This new sample will contain about 9% (30% of 30%) dead units. Thus, it will no longer be representative of the population of dead enterprises on the SBR, which is still nearly 30% assuming that the survey sample is a relatively small proportion of the population. There are proportionally too many live enterprises in the sample. If the weighting procedures do not take this into account (by allowing for the dead enterprises that were found in the sample) the result will be an upward bias in the estimates. Furthermore, the bias will get increasingly worse with each survey repetition.

Register updating information that cannot be obtained from the administrative source on which the SBR is based, or from survey feedback, has to be obtained using register improvement surveys and re-profiling operations conducted by the business register staff. These are resource intensive activities but essential, for example when the owning institutional units of large enterprises go through complicated changes like mergers, amalgamations, split-offs, etc.

6 Statistical Business Register Functions

6.1 Provision of Survey Frames.

The primary and most important function of the SBR is to provide frames for list based enterprise surveys. Given the particular set of enterprises (or other producing units) in scope for a particular survey occasion, it must be possible to generate the corresponding frame. For example, for an annual survey of manufacturing, for reference year Y, mailed out in, say April Y+1, it must be possible to extract a frame in, say February Y+1, containing all the manufacturing enterprises that were active in year Y, and containing the latest contact data, i.e. as of February Y+1.

A measure of the success of the SBR is the extent to which it is actually used. Evidently, a very well designed SBR is of no practical value unless it is the source of survey frames.

In addition to provision of survey frames, the SBR has other related functions as outlined in the following paragraphs. They are supplementary to the main function and should not be allowed to detract from it.

6.2 Sample Selection and Sample Overlap Control.

Given that the SBR is the source of enterprise survey frames, it is vital that the SBR frame extraction software interfaces nicely with the sample stratification and selection programs and initiates the survey input database in which the survey data will be stored. It is also important to ensure that small enterprises do not find themselves included in a large number of different surveys. This can be achieved through sample overlap control facilities. There are various different methods in current use, including assignment of a permanent random number to each statistical unit and selection of non-overlapping samples based on these numbers (Australia and Sweden provide examples) and selection of samples with probability inversely related to the number of surveys which the units are already in (Netherlands and France provide examples).

6.3 Respondent Burden Monitoring.

It is very useful to be able to list all the surveys to which any given enterprise has to report each year. It is also important to be able to measure the respondent burden imposed on enterprises by individual surveys and by all surveys in total. These functions are most efficiently handled through the SBR.

6.4 Production of Statistics and Links to Other Data

Cross-sectional statistics. There is inevitably a requirement to be able to produce counts of enterprises (or other statistical units) in response to ad hoc requests.

Statistics on business dynamics. A facility for producing data on enterprise demographics – births, deaths, changes over time – is also very useful.

Intercorporate ownership and control data. Depending upon the ownership and control information stored in the SBR there may be a facility to produce ownership trees and tables.

Links to additional data in other databases. The SBR may be a tool for linking data about an enterprise across databases.

6.5 Frames for External Surveys of Enterprises.

A common request from other government agencies and even private organisations is for lists of enterprises to provide frames for their surveys. Whether or not this is possible depends upon the legislation under which the statistical office operates. Usually there is a clause which guarantees confidentiality of statistical information and this may prevent information

about individual enterprises being given out. Also, some of the data in the SBR may have been derived from an administrative procedure that is itself subject to a restriction on distribution. On the other hand, much of the data about enterprises stored in the SBR may, in any case, be publicly available. Several national statistical offices operate under legislation that specifically allows the dissemination of the name, address, principal economic activity and size code of enterprises, deeming these to be public knowledge.

6.6 Administrative Business Register.

In the case where the register is based on legislation that is specifically designed for statistical purposes and/or is administered by the national statistical office, a single register system may also serve administrative as well as statistical functions. However, these administrative functions should be viewed separately as they involve different needs. In particular, they include provision of data about individual enterprises on request. This may not be satisfiable within the framework of the legislation governing statistical data because of confidentiality provisions. Also, administrative requirements imply the need to update the administrative register with those and only those data that come through the administrative processes.

7. Statistical Business Register Organisation, Procedures and Systems

7.1 Organisational Framework and Human Resources

The overall organisation of the statistical office and the way in which economic statistics program is designed and conducted influence the role to be played by the business register. Conversely the quality of the business register affects the design and conduct of program. Thus, office organisation and program design are factors that need to be taken into account in determining the appropriate business register structure and that thus influence coverage.

There is no "right" organisational structure for a SBR. Staff can be in one location or they can be spread across the regions. Their organisation depends substantially upon overall organisational structure of the statistical office.

- For example, Statistics Canada is strongly centralised, having relatively small regional offices whose main focus are personal interviews and dealing with data users. In this case the SBR and staff are located at the Ottawa central office. In France, there are staff responsible for collecting information for the register throughout the country, but the register is controlled from the INSEE central office in Paris. Both organisations may be said to be operating with a *large head office model*.
- In Australia, each regional office has a different range of responsibilities that it performs on behalf of the whole organisation. The SBR and staff are concentrated in a single location, but in the Melbourne regional office not the central office. The ABS may be said to be operating on a *distributed head office model*.
- In Germany, on the other hand, large regional offices all have identical functions, including all types of data collection within the region. A SBR is maintained by each regional statistical office and the data from these have to be combined to form a notional national register. The German Statistical Office may be said to be operating on a *regional office model*.

7.2 Procedures and Systems

All procedures must be documented so that staff responsible for maintenance and users know what is expected of them.

The SBR systems should satisfy the functionality requirements for creation, maintenance and use described above. Ideally there should be a single database to which all users have access

and updating rights according to their needs through their desktops. However, there is no "right system". It will be dependent upon the information technology available and the overall systems architecture in the statistical office. For example, the system may comprise a central database to which all users have access through a network, or it may comprise a set of regional databases that are synchronised at regular intervals.

8. Coverage Related Quality and Performance Indicators

8.1 Introductory Remarks

An essential feature of the operations of an economic survey programme is a quality management and measurement framework. Quality measures should be defined and measurements made on a regular basis. This section lists quality and performance indicators that are directly or indirectly related to coverage. They are presented in the form of questions under five headings: conceptual framework, collection strategy, frame creation, SBR functionality and usage, and management.

8.2 Conceptual Framework

If the conceptual framework is not well defined, coverage itself cannot be well defined or measured.

- Is there a well defined statistical units model, in accordance with international standards (in particular SNA93), providing appropriate breakdown of economic activity without being overly complex?
- Are the classification systems appropriate?
  - *international standards for classification by economic activity and by institutional sector;*
  - *international/national standard for classification by region;*
  - *national standards for classification by size, type of ownership/legal form, and activity status;*
  - *well defined classification coding instructions.*

8.3 Appropriate Choice of Collection Strategy

Collection strategy has a profound effect on coverage

- Have administrative data sources been thoroughly examined to determine to what extent the data they contain can be used to support the statistical programme? Are there unused administrative sources that would help address coverage problems?
- Is there any way in which the present set of surveys should be combined or split to provide more effective coverage?
- Is there an alternative survey design that would give better coverage, for example use of mixed household enterprise survey in place of enterprise survey for small businesses?
- Are there any surveys in the present programme that can be eliminated totally, thus releasing resources that could be spent in improving the quality (including coverage) of other surveys?

8.4 Frame Creation and Maintenance

Where a survey is using the business register these indicators refer to the business register.

- Is effective use made of administrative sources for recording new economic entities?
  - *Harmonised use of all pertinent sources;*
  - *Maximal use of the registration and transaction data available from each source.*
- Are there well defined effective maintenance procedures?
  - *Full use of administrative data available;*
  - *Appropriate use of feedback from surveys;*
  - *Register improvement surveys;*
  - *Profiling procedures for handling all forms of structural change in large enterprises.*
- What coverage is provided by relative to the conceptual target population, for example as defined in SNA93?
  - *Type, number and percentage (by number and employment/sales) of units excluded by design;*
  - *Type, number and percentage (by number and employment/sales) of units excluded as result of their failure to register when they should;*
  - *Type, number and percentage (by number and employment/sales) of units excluded as result of the failure to carry their registration through to SBR;*
  - *Type, number and percentage of units (by number and employment/sales) excluded because of the time lag between registration and recording in SBR.*
- Are the activity status data correct?
  - *Number and percentage of inactive or dead units flagged as active;*
  - *Number and percentage of active units flagged as inactive or dead.*
- Are the contact data correct?
  - *Number and percentage of units with missing contact information;*
  - *Number and percentage of units with wrong contact information.*
- Are the sector classification data correct?
  - *Numbers and percentage of units with missing sector classification;*
  - *Numbers and percentage of units with wrong sector classification.*
- Are the primary activity data correct?
  - *Number and percentage of units with missing activity classification;*
  - *Number and percentage of units with wrong activity classification;*
- Are the size classification data correct?
  - *Numbers and percentage of units with missing size classification;*
  - *Numbers and percentage of units with wrong size classification.*
- Are the geographical classification data correct?
  - *Numbers and percentage of units with missing geographical classification;*
  - *Numbers and percentage of units with wrong geographical classification.*

8.5 Statistical Business Register Functionality and Usage

The following measures apply to the business register and indicate its effectiveness in providing survey coverage.

- Is the SBR fully utilised as the source of enterprise survey frames
  - *Number of surveys for which survey frame derived solely from SBR.*
  - *Number of surveys for which survey frame derived from SBR but some corresponding frame data items are obtained from elsewhere.*
  - *Number of surveys for which survey frame is obtained from elsewhere, supplemented by new enterprises from the SBR.*
  - *Number of surveys for which survey frame is obtained from elsewhere without any reference to SBR.*
- Are there effective and efficient survey frame generation procedures?
  - *Facilities to readily identify the statistical units belonging to any specified survey population;*
  - *Facilities to generate a survey frame for any specified reference period.*
- Are there sample selection and monitoring procedures?
  - *Facilities enabling control of sample overlap across surveys;*
  - *Facilities enabling monitoring of respondent burden.*
- Are there data analysis facilities?
  - *Enabling snapshots of the populations of statistical units;*
  - *Enabling analysis of business dynamics.*

8.6 Business Register Management

The following measures apply to the business register management and reflect indirectly on its effectiveness in providing survey coverage.

- Is there a quality measurement and management program?
  - *Quality measures are defined;*
  - *Measurements of data quality are made*
  - *There is quality awareness and orientation.*
- Are human resources adequate and well organised?
  - *Adequate numbers and skills of staff;*
  - *Appropriate distribution of staff across the central office and regions*
- Is documentation adequate?
  - *comprehensive manuals and reference materials;*
  - *easy access to documentation;*
- Are computer systems well designed and functional?
  - *Appropriate systems architecture;*
  - *Ease of access to, and update of, data;*
  - *Ease of survey frame extraction and report generation.*

Part C: Other Types of Survey and Coverage Issues

9. Other Types of Survey

9.1 Introductory Remark

Whilst list based enterprise surveys are the most commonly used, there are other types of surveys collecting economic data for which coverage issues need to be considered, as outlined in this section.

9.2 Area-Based Enterprise Surveys

In an area based enterprise survey, the initial sampling units are a set of geographical areas. After one or more stages of selection, a sample of areas is identified within which enterprises are directly listed. From this list, the sample is drawn and data obtained. As previously noted, this approach is less efficient than a list based approach.

9.3 Household Surveys

Household surveys are widely used for collection of social statistics. In addition they are valuable in providing coverage of economic production by household enterprises that are too small to be recorded in any readily usable administrative list of enterprises. As household surveys exist for the purposes of collecting labour force and household expenditure data, additional questions related to production activities can be added at relatively little extra cost. This makes the use of a household survey generally cheaper than conducting an area based enterprise survey for the same purpose. However, the responding unit is a person in a household, not an enterprise, thus the data that can be collected about the activities of the enterprise may be correspondingly more limited.

9.4 Mixed Household-Enterprise Surveys

In a mixed household-enterprise survey, a sample of households is selected and each household is asked whether any of its members is an *entrepreneur, i.e.*, the sole proprietor of, or a partner in, an unincorporated enterprise. Data for the enterprises thereby identified (or for a sub-sample of them) are then collected from the respondent reporting on behalf of the enterprise, either immediately or in a subsequent stage of data collection. Thus the feature of a mixed household-enterprise survey that distinguishes it from a household survey is that it collects information about enterprises per se, whereas a household survey collects information about the persons in a household, including possibly their personal contributions to enterprises.

Mixed household-enterprise surveys can thus provide coverage of small enterprises that are not included in list based enterprise surveys. However, they suffer from similar disadvantages to area based enterprise surveys, namely the inefficiency of the sample design and the difficulty of handling enterprises with production units in more than one location.

In addition, data for an enterprise that is a partnership may be reported by each of its partners who may be in different households. The duplication of coverage that this implies has to be allowed for in the survey estimation system. This is the feature that distinguishes a mixed household-enterprise survey from an area based enterprise survey, as, in the latter case, enterprises are directly identified and listed (hopefully) without duplication. The process of producing an unduplicated list is the reason why area based enterprise surveys are generally more expensive than mixed household-enterprise surveys.

In summary, mixed household-enterprise surveys are sometimes preferred to household surveys or area based enterprise surveys for estimating the production of small units that are excluded from list based enterprise surveys.

9.5 Indirect Enterprise Surveys

An example of an indirect survey is where the enterprises that administer city markets are asked for data about the holders of the market stalls. This sort of survey provides only limited data about the observation units and often only in aggregate form. For these surveys coverage has two aspects:

- coverage of the reporting enterprises by the statistical office;
- coverage of the target enterprises by the reporting enterprises.

9.6 Price Surveys

Producer and consumer prices are usually collected by entirely separate surveys from those used to measure production or expenditure. With few exceptions, the survey samples are not probability samples – the items selected for pricing and the enterprises from which the prices are collected are chosen purposively. Thus representative coverage not exhaustive coverage is the goal.

10. Area Frames for Household, Enterprise and Mixed Household-Enterprise Surveys

For household surveys, the direct equivalent of the business register is a household register. However, in most countries there is no administrative source on the basis of which a household register can be readily constructed and maintained. Thus an area frame is commonly used as the starting point for household surveys, in particular for labour force and household budget surveys that are likely to form part of the basic programme. The quality of survey coverage is largely determined by the quality of the area frame.

Construction and maintenance of a household area frame involves:

- division of the country into area segments, using information about the numbers of households in each segment obtained from the population and housing census;
- selection of a representative sample of segments;
- a two, three, or even four stage design – typically involving different treatment of urban and rural areas – the penultimate stage being enumeration of all households within the areas selected and the final stage being selection of a sample of these households;
- systematic maintenance of the selected areas and enumerated households;
- replacement of the frame following the next census when new information on the numbers of households in each area is available.

Similar design principles are used for area based enterprise surveys as for area based household surveys. However, as the ultimate object is to enumerate enterprises, the size criterion used in delineating the areas is the number of enterprises rather than the number of households.

11. Web Based Surveys and Coverage Issues

A web based survey is computer assisted, self-administered survey in which some or all questionnaires may be completed on line and the data transmitted directly via the internet to a database belonging to the data collection agency.

A web based questionnaire has some similarities and some differences from a paper questionnaire.

- A more limited range of design elements is available in HTML than on a paper version. The range of design elements may be supplemented by use of Javascript and there are opportunities for use of a wide range of images.
- There are opportunities for interaction with the respondent as the questionnaire is being completed.
- Security is considered more of an issue, though being addressed through use of digital certificates.
- The different collection mode may result in systematically different responses from the same respondents.
- Data collection and processing costs are much less.
- Use depends upon the respondents' IT facilities.

So far as national statistical institutes are concerned, a web based questionnaire is viewed as an option in the context of multi-mode data collection. No NSI is attempting to conduct a mainstream survey solely using the web as not all respondents have internet access or necessarily want to use it in replying to surveys. Web based questionnaires are simply another collection medium like paper, telephone, and fax. Thus in this context, there are no particular coverage issues.

References

Chapters 2-6 of Business Survey Methods, Wiley

Eurostat Handbook on Business Registers and Survey Frames

UNECE Business Register Questionnaire 2004 and Guidelines for Updating (Business Register) Identity Card