

FOR PARTICIPANTS ONLY

REFERENCE DOCUMENT

DDR/7

3 April 2001

ENGLISH

ORIGINAL: SPANISH

ECLAC

Economic Commission for Latin America and the Caribbean

First meeting of the Statistical Conference of the Americas of the
Economic Commission for Latin America and the Caribbean

Santiago, Chile 9-11 May 2001

COMPANY REGISTRIES FOR STATISTICAL PURPOSES

Their role in production statistics

This document was prepared by the National Institute of Statistics of Spain. It has been reproduced without formal editing.

01-2-175

Aim of report

This report discusses the main reference documents to be considered when compiling and maintaining company registries for statistical purposes.

The availability in national statistical offices of an operational and up-to-date system of information on the set of firms operating in the country, is considered internationally to be essential for carrying out necessary statistical operations. The veracity, relevance and internal consistency of statistical data on business activity, are likely to depend largely on the degree of development and functionality of the company register.

This document is organized as follows:

The first section provides an overview of the most important aspects of business statistics.

The second section provides a general description of how a harmonized system of business statistics began to be constructed in the European Union, highlighting the crucial role played by infrastructure.

The third section describes the main methodological guidelines that have been used in developing or administering business registers nationally, in order to adapt them to community standards.

The fourth section outlines the key work done on this topic by the National Statistical Office of Spain, identifying the main lines of action and briefly describing the final output obtained.

The fifth and final section proposes some activities to be undertaken in the short or medium term in the field of company registers, setting out the underlying motives.

1. Business statistics: specific aspects. The heterogeneity of the enterprise universe

Official statistics programmes traditionally distinguish between statistics on households and those focusing on enterprises. This separation is clearly shown in the organizational structures of nearly all national statistical offices. Nonetheless, it is an open question whether this organizational structure reflects any fundamental difference, i.e. whether business statistics have specific characteristics that distinguish them from statistics on other topics. The general opinion is that specific reference elements can be established in relation to the population of enterprises operating in a given geographic domain.

As first sight, preparing business statistics might appear to be a very simple task: all that needs to be done is to compile a register of enterprises, with certain quality assurances, and then question a representative sample about their main accounting figures. The later stages of the process would consist of compiling the results and then carrying out the appropriate integration or aggregation, using suitable estimation formulas.

This initial idea collapses immediately, however, as soon as one sets to work. Anyone who has had responsibility for this type of investigation, will have discovered that every phase of the process is actually highly complex, for three basic reasons:

- The extraordinary heterogeneity of the universe of businesses in existence;
- Problems relating to data collection; and, especially,
- The increasing information demands that businesses have to respond to.

The key characteristic of a population of enterprises, compared to other populations, is its heterogeneity. The fact that this makes the issue more complex ought to act as a stimulus to the researcher, since heterogeneity is the underlying motive for statistical analysis. So, where is the problem?

The difficulty essentially arises when quantifying this heterogeneity. Enterprise size, for example, can be expressed through sales figures or the number of people employed. These sizes imply weightings for each unit and, consequently, different impacts on the final aggregates. In practice there is huge dispersion in these weightings. Large auto manufacturers or oil refineries are registered as a single units in the corporate register, just like the neighbourhood greengrocer. Yet their contribution to the final data may be extremely important, and in some cases the absence of a certain unit could seriously distort the information or even completely invalidate it.

This heterogeneity calls for highly specific treatment of informing units and in the way samples are defined (exhaustive strata are indispensable), together with efforts to ensure that data provided by firms are of good quality. In this regard, very large firms have complex organizational structures and may be sensitive about data confidentiality. In their favour, however, they generally have quite robust internal information systems. At the other end of the scale, the smallest firms tend to have very unsophisticated information systems that are sometimes inadequate for satisfying statistical information requirements. This is one of the many examples that leads to individualized treatment of different subpopulations of business units.

In addition, firms have to support increasing demand for annual, quarterly or even monthly information from a variety of public administration bodies, on which compliance is normally compulsory. Such demands mean increasing costs for the informing units, particularly those in certain sectors or size categories that inevitably attract systematic research. All of this requires special care when designing

enterprise data-capture operations, to take into account the potential effect such practices may have on data quality.

2. Construction of the European economic statistics system. Infrastructure regulations

The 1990s have seen a major drive to renovate national systems of business statistics within the European Union. During the first half of the decade, the Statistical Office of the European Communities (EUROSTAT), launched a set of initiatives to develop comparable statistics in all member States.

Supranational data comparability involves two aspects of harmonization. The first relates to the availability of statistical data on a given phenomenon and for a specific time period. The second extends this with the additional requirement that concepts, definitions, data collection and handling procedures, and, in particular, the methodological base that underpins the statistical data, must admit relevant comparisons between the different national aggregates.

Based on these principles, EUROSTAT has made an exhaustive analysis of the availability and methodology of statistical data in member States, giving priority attention to the following aspects:

- Coverage of data available
- Units used for supplying information
- Time reference of the data
- Data classification

The main conclusion to emerge from this analysis was that both the scope and the nature of the data available in different countries was extremely varied. From the coverage point of view, the services sector was a great unknown. The units of observation and analysis used in statistical surveys varied from country to country; and even where there were common units, the same conceptual criteria were not used. The time reference of the final data and the period for which they were available were also not harmonized, nor were systems for classifying and stratifying the information.

In view of these general shortcomings, in order to successfully implement the reform, EUROSTAT is organizing a comprehensive project aimed primarily at creating and developing a genuine common statistical infrastructure, embracing concepts, nomenclatures, working methods and tools. This whole project is based on the following three European Council Regulations:

NACE Rev.1 (Regulation No. 3037/90)

This regulation establishes a General Industrial Classification of Economic Activities in the European Community with the specific aim of classifying the production processes of enterprises and other statistical units.

On a worldwide scale, an obligatory reference for the classification of economic activities is the United Nations International Standard Industrial Classification of All Economic Activities (ISIC). This classification is conceived as an international framework for comparison, suitable for handling data from developed countries and developing economies alike.

It is logical, therefore, that the European Union should develop a classification system that adheres to the fundamental principles of ISIC, while also faithfully reflecting the greater diversity and economic development of EU countries.

It was decided that the third revision of ISIC (ISIC-Rev. 3) would form the basis for NACE-Rev.1, and that the first two levels of each classification would be identical. In the third and fourth levels of ISIC-Rev 3, product lines would be disaggregated where necessary, but always guaranteeing the exact reconstruction of a subdivided ISIC product line from different NACE categories.

Regulation on statistical units (No. 696/93)

The concept of statistical unit is quite difficult to define in principle, since it is often called by different names such as “declaring unit,” “informing unit”, “observation units,” “unit of analysis”. The regulation refers to all of these as “statistical units of the production system,” highlighting their faculty to undertake economic activities, and thus distinguishing them from other statistical units.

This regulation lists and defines eight types of statistical unit that need to be considered when preparing economic statistics. It requires member States to use exclusively those units and the corresponding terminology in the development and execution of statistical operations, and also in the official publication of results and transmission of data to EUROSTAT.

Regulation on business registers for statistical uses (No. 2186/93)

This regulation obliges member States to set up and maintain an information system that covers the set of enterprises and local units that exist in the country, with certain requirements concerning coverage, units and their associated characteristics, and the updating of data.

This is closely related to the two previous regulations. Firstly, the units to be registered need to be adapted to the definitions set out in the regulation on statistical units. In addition, the most important variable for classifying a unit is its main economic activity, and NACE Rev.1 is explicitly referred to as the only valid instrument for codifying the business register population.

In order to apply these regulations effectively, the procedures used have adhered to working methods universally accepted by the European Union. The work was preceded by several years of preliminary analysis, consultations with users and data providers, and consensus reached with the various national statistical offices. This whole effort is now starting to bear fruit.

3. Guidelines for the administration of company registries

As part of the programme to harmonize and develop business registers, the countries of the European Union set up a permanent working group with the aim, among others, of drawing up a set of methodological standards to serve as reference when promoting or carrying out statistical work at national level. An outline of the most important guidelines is provided below.

3.1 OBJECTIVES, COVERAGE AND CONTENT OF A REGISTER

In the first place, it is essential to analyse and clearly define the main objectives and potential uses of the registers, taking their foreseeable evolution into account, to determine the reference elements that need to be considered in their implementation and subsequent development.

The priority function of a register is to provide the reference instrument for undertaking and coordinating economic surveys. Its role as a generating framework should be understood in a broad sense and includes:

- The initial availability of the subpopulations under analysis, together with the basic magnitudes associated with them.

- Application of the usual sample selection techniques, generating individualized files of units for research in the field.
- The contribution of raising factors and any other magnitude enabling population results to be inferred from sample data.

Just as important as survey execution is coordination. If the register is central and covers all sectors of activity under analysis, it will be possible to make sure each firm receives a single questionnaire in production surveys. Correct use of the register will also make it possible to control the statistical burden that enterprises support and take measures to ensure that small firms, at least, are not repeatedly selected.

Aside from this main objective, other important aims can also be considered, and in fact the register may prove to be a very valuable source of statistical information. It is an observable fact that the demand for statistical data on the population of enterprises operating in a given country is growing, and new information needs are emerging all the time. Not only is it interesting to know how the business fabric is distributed at a given point in time, but data are also needed showing the evolution of the various populations over time. Information on the number of business start-ups or closures, with corresponding breakdown by activity and size, is widely demanded by different users. Data on phenomena such as mergers, takeovers or corporate breakups should also be covered by the register, although perhaps not as a priority objective in the initial phases of implementation.

As specified in the corresponding community regulation, the business register should encompass the set of enterprise units contributing to each country's gross domestic product.

Certain units can be excluded, however, if they are economically insignificant, or else for cost/benefit reasons. Others are optional depending on the needs of each country.

Units carrying on activities in the sectors of agriculture, hunting and forestry (NACE Rev.1, section A), fishing (section B), public administration and defence: compulsory social security (section L), have been declared optional. Units engaged in activities under sections P (domestic services) and Q (activities of extra-territorial organizations) are outside the register's field of observation. Inclusion of the remaining economic activities is mandatory.

Coverage can also be analysed from the standpoint of size. The increasing interest being shown in the activities of small and medium-sized firms makes it necessary to include firms of all sizes. When formulating this objective, however, one should not underestimate the major difficulties entailed in fulfilling it, particularly in the case of small firms.

Certain tax administration files may be a good source of information in this respect. In particular, files relating to local taxes levied on economic activities may provide reliable sources for detecting small enterprises that fall outside other fiscal domains such as VAT. Moreover, the vast majority of enterprises that perform professional services are members of their corresponding professional associations, which provides additional ways of detecting and identifying them.

Lastly, we refer to the main variables that need to be associated with business units in the register, bearing in mind that these must be compatible with the uses or purposes envisaged for them. The data to be recorded for each enterprise can be classified as follows:

(A) Identification variables

Identifier

The essential quality of the identifier is that it should remain unchanged throughout the entire life of the enterprise. Experience shows that a good solution is to assign each unit a non-significant sequential numerical code and give it an information control key. If the administration of the country generates a unique number for all enterprises, regardless of the information system in which they are registered, then statistical services are recommended to use that code in their processes.

Name

It is useful to distinguish between natural and legal persons.

In the first case this means precisely recording the name and surnames of the individual entrepreneur concerned, and, in the second case, the official registered name of the enterprise.

Legal form

This information may be very useful as a possible selection or stratification criterion to be used in surveys. Experience shows that it is normally helpful to adapt questionnaires and data collection processes according to the legal form of the enterprise. In this variable, natural persons will need to be distinguished from legal persons, and, within the latter, the different categories that each country establishes in its national legislation.

Postal address

Every enterprise will have (at least one) postal address assigned to it, corresponding to its registered office. This address should be associated with the corresponding geographic code according to the Nomenclature of Territorial Units for Statistics (NUTS). As a minimum, level five (municipal) of this classification is compulsory. Addresses will be registered in full, wherever possible, with all the specific indications relating to current postal norms (postcode, postbox, etc.), and a business contact number (telephone, fax, electronic mail, etc.)

(B) Classification variables

Economic activity

In order to define suitable sampling schemes and prepare economic statistics on a consistent basis, statistical units should be classified by their main activity according to NACE Rev.1 at the 4-digit level. If an enterprise carries on several activities, the main activity will be decided on the basis of value-added by each activity undertaken by the unit. If this information is not available, alternative criteria could be used such as employment or sales volume.

As regards other activities of the enterprise, if data exist that make it possible to identify significant secondary activities, i.e. if they involve at least 10% of the total activities or gross value added at factor cost, or if they represent 5% or more of this type of activity nationally, the record for the unit concerned should contain a reference to the secondary activities.

Size

Several criteria can be used to construct a size indicator, the most important being persons occupied and sales volume. The size of units is crucial for sampling schemes and the extrapolation of results.

(a) Measure of size by labour force

The ideal measure here is the number of persons occupied. Enterprises are not obliged to have a numerical figure available, but a size class needs to be assigned. The best time frame is to register average annual employment rather than the exact employment figure on a particular date.

(b) Measure of size by turnover

In some samples drawn from very specific activity sectors, it is not appropriate to select the sample on the basis of employment. Other indicators are needed such as sales turnover, taken exactly as this appears in the firm's accounts. In the event of the exact number not being known, each enterprise must be allocated to a size class defined in millions of Euro.

(C) Operational variables

Activity status or situation

Top priority should be given to dealing with active enterprises, but those that have ceased activities also need to be considered for various reasons. This is relevant for sample selection aimed at studying the main characteristics of this population, and also because such enterprises might start up again some time in the future and would have to be included once again. Consequently, it is important to have a code specifying the status of the unit as “active” or “inactive”.

Date of commencement (cessation) and of entry in (removal from) the record

Associated with the status code, the date of the unit's first inclusion in the register will be recorded, along with the date on which it commenced activities. If the unit is inactive, the date of cessation of activities should be recorded along with the date of removal from the register.

These are the main variables that need to be operational for each unit in the register. Obviously these can be complemented with other relevant content, such as indicators on participation in surveys, control links to other units, the public/private nature of the unit, or any other content deemed relevant for potential users.

3.2 STATISTICAL UNITS OF THE REGISTER. DEFINITION OF THE ENTERPRISE

The Regulation on business registers envisages registration of three different types of unit, as follows:

- Enterprises

- The *legal units* responsible for them
- The *local units* that depend on them

The conceptual model thus makes the enterprise the basic and central unit of the register, to which other types of units, namely legal and local units, are related. There is also explicit reference to the Regulation on statistical units, since this is the legal framework defining such units.

Firstly we analyse the concept of the legal unit. The regulation specifies that, in principle, legal units are not considered as statistical units, as they do not always have sufficient autonomy to carry out a productive activity. Nonetheless, their identification is crucial for constituting the statistical unit of the enterprise. There are two categories of legal units, namely:

- *Legal persons*, whose existence is recognized by law independently of the individuals or institutions which may own them or are members of them.

- *Natural persons*, who are engaged in an economic activity in their own right.

The legal unit is thus a unit endowed with legal status, i.e. the ruling legal framework assigns it a series of rights and obligations. All countries recognize rights for both natural and legal persons, including ownership, entering into contracts with third parties, and recourse to the justice system to defend their interests.

It is clear that legislation can only grant rights to, or impose obligations on legal units. This is one of the reasons why statistical services have to take them into consideration, since it is these units that are subject to the obligations established in national legislation on public statistics.

As indicated above, however, the legal unit sometimes may not be the relevant unit for understanding productive processes. There are circumstances where several legal units combine to carry on an economic activity jointly. The relations or ties that exist between these units have to be taken into account for a global understanding of the production process concerned, and thus for quantifying inputs, outputs and, especially, the operating surplus.

The statistical definition of the enterprise is based on these premises. The regulation states, “the enterprise is the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources.” As discussed below, an enterprise may be a sole legal unit. In practice, this is the most frequent case.

As can be seen, the definition is somewhat complex, and this causes problems of understanding (even among statistics professionals), together with problems of operation or practical demarcation. The key task that still remains to be fully resolved is to identify the set of situations for which the legal unit does not have *sufficient decision-making autonomy* in carrying out a productive process, and, thus has to be combined with other legal units to constitute the enterprise.

Work done on this issue in the European Union has identified three elementary types of business combination:

The simple enterprise

As mentioned above, this is the most general case. A sole legal unit is the legal basis of the enterprise as a statistical unit. Consequently, the two units are essentially the same.

Joint venture

This situation occurs when several legal units jointly carry out an activity without strong ties existing between them. None of them owns or controls the others. Generally speaking, the relations created between such legal units tend to be informal or contractual, and usually such ties give rise to a new entity, which, while not endowed with its own legal status, is nonetheless treated as such under fiscal or social legislation. The entity generated in this case is known as a quasi-company, although it goes by specific names in the legislation of each country.

In this case, not all the legal units associated with the quasi-company have to be registered, provided they do not carry on any other activity independently on their own account. This case would reduce to a single enterprise with a sole legal basis rather than several equivalent legal units.

The complex enterprise

This is definitely the most difficult case to identify and deal with in practice. A necessary condition for making an appropriate combination of legal units is prior detection of a control link running from one legal unit to another or others. What criteria should be followed from this point?

The correct procedure is indicated in the definition of the enterprise itself, specifically in these two expressions: “it is an organizational unit producing goods or services”, and it “benefits from some degree of autonomy in decision-making.” This therefore involves combining legal units belonging to a group, which at the same time jointly form units within it with autonomy to carry on production processes. Thus far, the theoretical basis of the process has been laid down; but how should one proceed in practice?

The explanatory note attached to the definition of the enterprise provides a clue: “... Some legal units, in fact, perform activities exclusively for other legal units and their existence can only be explained by administrative factors without them being of any economic significance [...]. In many cases, the activities of these legal units should be seen as ancillary activities of the parent legal unit they serve, to which they belong and to which they must be attached to form an enterprise used for economic analysis.”

Ultimately, the enterprise as “an organizational unit producing goods or services” must have the capacity to combine factors of production, and it will be necessary for this combination to be complete for it to be able to produce goods or services. Therefore, if a legal unit has created other related legal units under its control to serve as a legal basis for certain factors of production, it will be advisable to consider this set of legal units as a single enterprise.

Such is the frequently the case of companies that set up a subsidiary for the sole purpose of acting as legal employer of part of the labour force needed to carry on the activity. Another characteristic example is provided by enterprises that set up legal units to “outsource” their ancillary activities (administration, accounting, informatics services, transport and distribution of production, ...). Ultimately, the range of sufficient conditions for the grouping may be highly varied and variable over time. The work that remains to be done on this is long and complicated, but little by little the foundations are being laid, and operating rules are being defined which are being adopted slowly but surely by statistical offices.

We conclude this section by briefly referring to the local unit for which, as mentioned above, registration is compulsory. This unit corresponds to an enterprise or part thereof, situated in a geographically identified place. At or from this place, economic activity is carried out for which (exceptions apart) one or more persons work for one and the same enterprise. The significance of this unit is that it more faithfully reflects the geographic location of activities, along with associated variables (billing, employment, investment, ...). Each local unit must be associated with the enterprise on which it depends, and this will make it possible to undertake comparative analyses between the two units.

4. Work done by INE: The Central Corporate Directory (DIRCE)

4.1 GENERAL DESCRIPTION

The situation developing internationally was compounded by internal shortcomings at INE, specifically the lack of a genuine central company registry, to serve as a valid reference framework for designing surveys and carrying them out.

Until the early 1990s, statistical research into one or more branches of activity meant compiling a specific directory. Construction of the appropriate reference framework thus made sense as a preliminary activity within the research process, but did not have its own independent rank or status.

This situation enabled surveys to be carried out (albeit with great difficulty in some cases), but the final product suffered from serious and multiple deficiencies. The most obvious problem was the complete lack of compatibility and internal consistency with the results of other operations in the statistical system, since the different frameworks were all developed independently.

Against this backdrop, in 1989 for the first time INE allocated resources to establish a set of ongoing tasks with a view to minimizing these problems. This is the starting point of a new project for integrating economic directories, generically known as PIDE (*Proyecto de Integración de Directorios Económicos*).

The initial activities of this project led to the conclusion that the model for generating and updating the register should be based on the appropriate use of administrative files. Subsequently, the system should be complemented and enhanced with information sources arising from traditional statistical operations or those designed for specific purposes. This point of view sees the business register as an endogenous element in statistical production. It is clearly the heart of the process and just as the register supplies data for research projects, it should in turn be fed back with all the data that such activities generate. This is the key to maximizing the internal consistency of national statistical systems.

Once the model had been established, the next step was to identify the optimal sources for supplying information. This phase of the project took into account aspects such as:

- Compatibility of different administrative sources with statistical aims (origin of data, territorial coverage, classification by activities and by size, basic units of the register, existence of content associated with those units ...)
- Quality of data in each source and degree of adaptation to reality.
- Access to the different files and how to overcome potential restrictions.

Periods of availability and the timeframe of the information

All these elements were analysed jointly for each potential information system. The main international experiences in this field were also taken into consideration, and two suitable data provision sources were identified, namely tax administration and social security

The activities of PIDE were thus directed at two independent subprojects: integration of the tax administration (PIDEAT) and social security (PIDESS) directories. Each of these in turn gave rise to an information subsystem, the central tax administration corporate directory (DIRCEAT), and the corresponding social security corporate directory (DIRCESS).

The tax administration directory PIDEAT was the first subproject to be developed, given the presence and quality of legal identifiers for each of the units. Based on previous studies and pilot experiences, including fieldwork, in 1989 an initial output system was constructed that was used as a temporary framework for incipient surveys focusing on the services sector.

In the case of PIDESS, following initial studies and relevant contacts with the government departments responsible, use was made of the available social security source best suited to the established goals, namely, the register of social security contribution accounts. This entry file was used for the first time in 1990, and the corresponding output system was generated.

These activities were later augmented by other processes that were fundamental for improving the coverage of the information system. Decentralization of tax management in the autonomous communities of the Basque region and Navarre caused an information vacuum when files obtained from traditional administrative sources were integrated. This gap centred mainly on the subpopulation of enterprises

without wage-earning employees engaged in activities in those geographic regions. Starting in 1993, two other subprojects were added, namely PIDEATN (directed specifically at Navarre) and PIDEV (aimed at the Basque region). In the case of Navarre, fiscal files were received, and DIRCEATN was created; for the Basque region, the process used the directory of economic activities administered by the Basque Statistical Office, to produce the DIRCEV system.

The resulting integration of the various statistical information subsystems, generated at each independent phase of the project, is known as the **INE Central Corporate Directory (DIRCE)**. The first edition of the register with national coverage took 1993 as its time reference. Since then there have been annual updates up to the present-day.

4.2 DEFINITION AND OBJECTIVES

The INE Central Corporate Directory (DIRCE) is thus the main output of the PIDE project.

It consists of an organized data set of magnetically stored information, with data including identification, location, territorial distribution and classification by size and economic activity of the enterprise and local statistical units. This is obtained from administrative sources or statistical files, and supported by other information obtained from ongoing INE statistical operations.

In the creation and maintenance of DIRCE, priority has been given to the general objectives established in EU methodological guidelines. In the first place, the aim of this statistical product is to make it possible to obtain subpopulations of units for use in surveys of enterprises or the local units where they carry on their activities.

Secondly, it is also a tool for generating statistical data on the enterprises operating in the country. Derived statistics could include the following types:

- Structural, quantifying the population of active units at a certain point in time, together with a description involving distributions according to geographic location, legal form, main activity or size class of the units concerned.
- Demographic, generating magnitudes that make it possible to study the evolution of businesses over time. This type of information is possible thanks to continuous updating of the system, which has made it possible to keep track of business start-ups and closures, together with their associated characteristics, for each update year.

DIRCE is also designed to satisfy demand for information on this topic from different international bodies, in particular EUROSTAT.

Lastly, one of the essential roles of the business register in a statistical office is as coordinator of the entire set of economic surveys. In this regard, the design of the DIRCE system makes it possible to measure the statistical information burden that firms support, and implement procedures to rationalize demand, especially for small and medium-sized firms which are obviously not obliged to participate exhaustively in the different statistical operations.

4.3 DATA ENTRY SOURCES

This section briefly describes the sources that supply data for DIRCE:

- **Tax administration sources** (the PIDEAT and PIDEATN subprojects)

Economic activities tax (IAE)

This tax is levied on every economic activity carried out by an enterprise in a given geographical location. The spectrum of production processes considered is very wide-ranging, covering business activities, together with those of independent professionals or even artists. Accordingly, this is one of the project's key files, as it includes very small firms, which usually represent one of the most widespread coverage problems in this type of work.

The basic record unit is the declaration of activity carried out by a enterprise in a specific place. One of the great advantages of this file is that it is a highly refined, and it is possible to construct enterprise and local statistical units from it, through groupings of elementary observations. In addition, each observation is accompanied by an indicator of the "status" of the activity declared, together with the date on which the activity commenced, and its date of cessation if this occurred during the reference year. These variables are fundamental for generating data on business demographics.

From the geographic point of view, IAE registration covers the entire State (all national territory apart from the Basque Region and Navarre). The same information also exists in a separate file for Navarre, which as mentioned above, manages its fiscal system autonomously.

Tax withheld on income from personal work (RTP)

This source comes from the annual summary of income taxes withheld on behalf of natural persons, and affects all enterprises that have satisfied income tax withholding requirements during the reference period. The declaration model involves an exhaustive report of income-earners and taxes withheld, although the information used is the global listing of persons from which a specific firm has withheld tax. This information is used as an indicator of the size of the unit, complementary to the rest of the information recorded in IAE. The geographic coverages available are identical to the previous file.

Value added tax (VAT) and corporation tax

Large firms, as measured by invoice volume, are characterized jointly by these two taxes. This source does not add new units to the register, since they have to appear in the IAE register anyway. They simply provide a size indicator for use in sample selection based on a criterion other than staff employed.

- **Social security sources** (PIDESS subproject)

Social security contribution accounts

The social security contributions account is an administrative management unit pertaining to the social security system. It encompasses a set of workers from a given firm that is homogeneous in terms of social security contribution, and whose labour activity is carried out within the same province. Strictly speaking, this cannot be considered as a local unit, although in many cases the two units are the same. This source provides data on identification, location, activity and size for firms with wage-earners located throughout national territory.

- **Statistical source** (PIDEV subproject)

Directory of economic activities of the Basque Region

The territorial decentralization currently in force in Spain gives rise to institutions with statistical competencies that are dependent on regional governments. The statistical office of the Basque Region includes in its brief maintenance of a directory of economic activities carried out within the region. Collaboration agreements established with the National Statistical Office include annual transfer of its information system for use in DIRCE updating.

The basic unit of the register is the local unit installed in the Autonomous Community, although it also contains information on parent firms, regardless of whether their central headquarters are located in or outside the Basque Region. With this data, all coverage problems in this geographic zone are overcome.

To conclude this section, it is important to stress that aside from these sources, other exercises currently in the development phase aim to make use of new administrative information which is likely to be relevant and indispensable in the near future. This relates to the data that firms are required to deposit in Commercial Registries, set up in every province in the country. The field of application basically pursues two objectives:

- Better knowledge of the demographic events that affect enterprises over time. These events may involve, on the one hand, the creation or disappearance of factors of production (business start-up or closure); and, on the other hand, their redistribution, through concentration (merger or takeover) or dispersal (partial or total breakup). The availability of information on the units involved in these events is crucial for demographic statistics derived from the register.

- Incorporation of the unit "enterprise group" into the information system. Definition of the control links between the different firms in the register is a key issue for statistical information on economic globalization. It will also make it possible to overcome shortcomings in statistical research into firms which, until now, has been based on the principle of national territoriality. This point is discussed below.

4.4 ECONOMIC, TERRITORIAL AND POPULATION DOMAINS

As is true of most registers of this type, administrative data sources have shortcomings that need to be overcome before processing the information, in steps prior to both implementation and any updating. This should be kept in mind when undertaking the activities needed to adapt the basic information to DIRCE requirements.

As provided in the community regulation on business registers, selection of the economic domain for DIRCE takes full advantage of data contained in entry files, applying appropriate filters essentially relating to legal status and economic activity.

Restrictions for the fiscal component are determined by exemptions granted in laws regulating tax administration: namely, public administration; social security administration bodies; public research and teaching organizations wholly financed out of public funds; and certain non-profit associations or foundations.

Legal status is mainly determined by the first letter of the legal identification code (NIF), present in all entry sources. Economic activity is indicated by the coding in each entry file, which is generally compatible with NACE.

Thus, each entry record is subject to a series of exclusion filters obtained by combining the initial NIF letter and activity codes that are not compulsory or outside the domain, as stipulated by the respective community regulation.

As a result of previous work, DIRCE covers all economic activities apart from agricultural and fishing production; central, autonomous and local government administration (including social security); activities of condominium owners and domestic service; and the activities of extra-territorial organizations. Private non-profit institutions that are common in certain fields (teaching, health, culture or social welfare) are only partly covered, as they are usually tax exempt. For these units, the priority data source comes from social security, which ensures registration of institutions employing one worker at least.

From the geographic viewpoint, the specified economic scope of DIRCE covers information for the whole of national territory. Lastly, under the population criterion, DIRCE records two different types of unit: the enterprise and the local unit. The construction and practical definition of these two units has adhered to the conceptual specifications established in the community regulation on statistical units.

The system does not impose any restriction on unit size and contains the population of enterprises with and without wage-earning employees, together with all their local units.

5. Action policies for the future

The consolidation and functionality of the business register should act as a stimulus to new activities in the short- or medium- term future. In this section, we suggest two lines of action that are currently highly topical and being discussed in all international forums.

5.1 REGISTER QUALITY PROGRAMMES

The ever-increasing importance of business registers in national statistical systems makes it essential for management mechanisms to implement procedures to measure and improve data quality. Such activities should go on alongside the processes of updating from administrative sources.

Generally speaking, the data contained in a register can be said to have quality when it satisfies three basic attributes:

- The data are relevant; i.e. the register uses statistical or administrative units that are appropriate for the intended purposes. In addition, the required variables for each of the units make it possible to correctly stratify populations for the purpose of selecting representative samples.
- The data are correct; i.e. the recorded information faithfully reflects reality; and
- The data are up-to-date; i.e. they relate to a recent time period.

The population of enterprises and their local units changes very quickly, and their characteristics may be in continuous flux. A good register should therefore provide the clearest possible picture relating to the most recent period possible. But these three elements of quality are not the only ones to consider. To ensure the consistency and coordination of national business statistics systems, the register should be *central* and *accessible*. Accessibility should apply to all statistical services, whatever the organization used in each country for collecting statistical data.

Maintaining high quality in a register is usually costly. Fortunately, however, a register does not have to be perfect to be useful as a statistical tool. It is therefore important to identify the main defects likely to be encountered and analyse their incidence and possible repercussions, in order to design procedures for minimizing them.

Not all defects in a record are equally significant. A list of the most important ones is presented below:

- Coverage deficiencies

Absence from the register of units that actually exist. Gaps of this sort may either be random or systematic in certain subpopulations of units.

- Duplication

The same unit may be registered several times under different identification numbers.

- Errors of identification, location and classification

This type of error is quite frequent among services. Classification errors in particular generate significant biases in estimations, increasing the variance of results and making it necessary to increase sample size and research costs.

- Falsely active units

Units that figure as active in the register, but which in fact ceased activity some time ago, increase the non-response rate and generate higher survey costs. This is one of the problems most frequently detected in registers, and it is inherited from the administrative files that supply the information.

There are several methods for quantifying quality defects, the most traditional being to run control surveys. This makes it possible to measure errors in classification by activity or size, and estimate the frequency of units that have ceased activity but still appear as active in the directory. On the other hand, the method does not make it possible to measure quality in terms of coverage defects and multiple units. For the first of these cases, alternative procedures include approximation using external files, the design of surveys by area, or controls based on employment, for example. In the case of multiple units, there are cross-matching techniques in informatics based on literal variables that generally give very good results. Nonetheless, the more accessible the register is to different users the more readily errors will be detected and corrected.

All these techniques are being implemented to a greater or lesser degree in European Union countries. For the future, however, procedures to obtain harmonized quality indicators need to be systemized and generalized, after which it will be necessary to intensify or re-target programmes on more important problems.

In any event, the unit responsible for the register must carry out an active improvement policy, remembering that quality is unlikely to be homogeneous throughout its entire domain. Activities relating to making updating sources more complete (mobilizing additional administrative sources or the results of production surveys themselves), the design of specific control surveys or opening up to potential users, should be included in daily processes of register management.

5.2 GLOBALIZATION AND ENTERPRISE GROUPS

Since the second half of the 1970s, there has been major evolution of the economy on a world scale. The development of international communications and information technologies, together with the growing liberalization of economic and financial exchange have accelerated the globalization of markets and enterprises.

From the business standpoint, there has been a sharp increase in events such as mergers, acquisitions or alliances, which are used by production units to expand internationally. A clear consequence of this is that the physical location of production is not relevant in the long run. Enterprises set up production centres based on the comparative advantages offered by different countries. Thus, a firm could perfectly well decide to carry out extraction activities in country A; manufacture intermediate products in country B, produce final goods in country C; do research and development in country D; and finally set up administrative and management services (accounting or marketing, for example) in country E.

In view of this, most statistical systems are currently based on the principle of economic territoriality. Data are gathered on the basis of operations carried out within the territory, and international transactions are usually integrated into national totals.

Governments have a large appetite for information on this phenomenon, since the development level in different countries and decisions taken in regard to it, should be based on detailed statistical description that makes it possible to obtain information on the strategies that firms adopt in organizing production, investments and commercial transactions, both within and outside national borders.

In late 1995, the EU Statistical Programme Committee (SPC) set up a discussion group with a mission to prepare a report evaluating the effects of globalization in the different statistical domains. The final report was presented to SPC in November 1998. One of its sections lists the statistical units involved in this globalization phenomenon. Apart from the statistical unit “*enterprise*”, which is defined at a national territorial level in the Regulation on statistical units, another basic unit of analysis is the “*enterprise group*”, which the same regulation defines at world level. This initial consideration is already becoming important in the management of business registers, since inclusion of this statistical unit is not compulsory under the Regulation on business registers.

There are major problems in defining enterprise groups, stemming from the different types of links that can exist between the individual firms in the group (control, ownership, considerable influence,...). Clearly, different relations can generate various definitions of the enterprise group unit, but the most widespread approach views the concept of controlling interest (total) as most appropriate, since this makes it possible to identify the group head uniquely and unambiguously.

Transactions among enterprise groups introduce significant biases and significantly reduce the quality of economic statistics as currently designed. It is therefore essential to adapt them progressively so as to reflect the reality and complexity of the world economy. The completeness of national statistics showing the involvement of resident units in the globalization process would be an important initial step. This supplementary information is needed both by producers of statistics and by the users thereof. Specifically, it would be important to have the following information:

- Classification of resident enterprises into foreign and locally controlled firms, together with their associated demographic statistics (e.g. number of units by activity and size, number of start-ups, closures, mergers, acquisitions).
- A supplementary classification of foreign controlled firms by country of installation and activity of group head.
- Weight of foreign controlled firms in national figures on employment, value-added, production, intermediate consumption, research and development, among others.
- Number of multinational groups under national control, together with their main characteristics (e.g. number of subsidiaries, installation, activity, employment, sales figures).

These examples clearly show that preparing such data is a complex and difficult task, especially when it is well known that firms are very reluctant to provide certain information about their internal structure and organization. Thus priorities have to be set, and more effective ways to gather this type of information need to be found.

The first important step is to incorporate the main enterprise groups into national data systems. This should be done through the national company registry, as this is the central tool for generating and detecting statistical units. For this purpose, it is useful to organize information exchange between the different sources at the national level and then, in a second phase, between the national statistical offices themselves, in order to optimize the use of existing data. National and international coordination is therefore particularly important in these tasks.

Final conclusions

The issues discussed in this report represent a compendium of the main methodological guidelines and specific experiences in operating business registers for statistical purposes.

Exercises carried out by the National Statistical Office of Spain, and in other European Union countries, are in response to the need to maintain a central business register to serve as a real driving force in national systems of statistics on production units, both in terms of practical execution and in terms of the compatibility and internal consistency of the final data.

To successfully set up and maintain the register, statistical offices must allocate the necessary resources in their budgets and permanently include a specific unit in their organizational structure with responsibilities for these activities and with sufficient rank to coordinate the system of business surveys.

For a variety of reasons, some countries base their survey frameworks exclusively on economic censuses. Generally speaking, this has serious disadvantages, including the high cost of these mass-scale operations and the static nature of the information obtained, which remains “as is” for a long time. A lapse of two years or more without updating is too long for the relevance of the information, when productive structures in real life change dramatically and almost continuously, at least among large enterprises. The problem for small firms is even worse since changes can be existential, and there is major turnover among populations.

Obviously there is no reason for accessibility, treatment and exploitation of administrative sources to be the panacea that solves all problems. Many files generated for non-statistical purposes have significant quality problems (e.g. file inflation resulting from undeclared cessation of activity, poor levels of coverage by activity or size, absence or poor quality of classification variables, units that are not compatible with statistical research).

Without doubt, one of the most complex issues in nearly all Latin American countries involves defining operating procedures to compile the set of production units classified within the so-called *informal* economy. Given their peculiar characteristics, such units are practically impossible to detect in administrative files. This issue, which has already been the subject of analysis and debate in a number of forums, would seem to require all countries involved to agree to analyse lines of action carried out and find a standard procedure to implement throughout the region.

In the general framework of compiling business registers, this problem may require a mixed model of action, involving continuous exploitation and annual updating of administrative data, combined with statistical research by areas (not necessarily exhaustive) every two years, for example.

In any event, one of the main conclusions to be drawn from an analysis of international experiences in this area, is that there is never a single administrative source that satisfies all the statistical requirements of a project. The files that are accessible or available all have their advantages and disadvantages, and professional activity should aim to strike a suitable balance between them in order to maximize efficiency and attempt to fulfil the objectives set out.