CES/SEM.46/WP.2
17 May 2001

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

COMMISSION OF THE EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE/EUROSTAT Seminar on Business Registers
(Geneva, 28-29 June 2001)

MODELLING THE ORGANISATION AND ECONOMIC ACTIVITIES OF BUSINESSES

Working paper prepared by Statistics Netherlands*

Keywords: classifications systems, statistical units, international harmonisation

INTRODUCTION

1. This paper describes a model of the structure and activities of businesses together with its main applications. The model has been developed as part of the European research project CLAMOUR (CLAssifications, MOdelling and Utilities Research), which belongs to the 5th Framework Programme. The contents of CLAMOUR as a whole is described by David Knight in another paper for this conference [1].

2. The purpose of the model is to provide a systematic, unambiguous description of the way businesses are structured, together with their activities. The description has to be detailed enough to allow two main applications:

---

* Prepared by Peter Struijs, Jacobiene van der Hoeven and Bert Kroese, Statistics Netherlands.

- The definition of statistical units in terms of the model. If data are collected on the basis of the model, statistical units can be derived by applying appropriate algorithms. As a minimum, the model is meant to support the statistical units defined in the Regulation on Statistical Units [2], but certain additional types of statistical units can also be defined in terms of the model.

- The specification of the classification system for the economic activities of the statistical units in terms of the model. If data are collected on the basis of the model, the statistical units can be coded according to the classification as well. As a minimum, the model is meant to support NACE Rev.1, but only for a very small range of economic activities. However, certain additional types of activity classifications can also be defined in terms of the model. Designing a new or updating an existing classification system for economic activities can also be done by reference to the model.

3. The model has to fulfil two main requirements. First, it has to reflect the needs of the users of statistical data concerning statistical units and activity classifications. In order to find out what these needs are, a large user needs survey has been included in CLAMOUR. The results have been an important input to the development of the model. The second requirement is that data collection based on the model must be feasible. This does not only mean that the respondents must be capable of providing the data, but also that the respondent burden must remain reasonable. In order to test the obtainability of information from businesses, Statistics Netherlands has developed a questionnaire, called the Data Providers' Questionnaire or DPQ. Details on the user needs survey and the DPQ can also be found in [1].

4. The two requirements work in opposite directions. Taking into account the user needs would result in an elaborate model, whereas data obtainability points in the direction of a rather simple model. In order to do justice to both requirements, the model has been designed in two stages: a comprehensive model has been developed based on the user needs, and a simplified model derived from the comprehensive model by taking into account the limitations of data collection. The relationship between the user needs, the data collection, the model and its uses is shown in Figure 1.

[Figure 1]

THE MODEL

5. The model developed in CLAMOUR is a static model. It is not aimed at following changes in the structure and activities of businesses through time. By necessity both the comprehensive and the simplified model include three parts:

- A part covering information to be collected (potentially).
- A part comprising other available information.
- A part representing the statistical units.

6. The first part concerns information on the structure and activities of businesses to be collected and used for the derivation of statistical units and classifications. The main data needs on structural aspects refer to the organisation structure and the accounting structure of businesses, together with the locations where activities are carried out, because these are key elements in the definition of statistical units. In addition, elementary data are needed on the activities carried out by the organisational units and at the different locations, allowing the classification of the various statistical units.

7. The second part is a consequence of the need to use as much as possible data that are available from administrative sources. These data mainly concern legal units and links between them, in particular ownership. In view of differences between the administrative systems of countries, the model has been kept fairly basic in this respect. Data on legal units are relevant in several respects. First, the main statistical unit, the enterprise, is defined in terms of these units. Second, when deriving statistical units from collected and administrative data, the universe within which the units are to be derived has to be specified. One of the ways to do this is by reference to the set of legal units that are part of the same enterprise group.

8. The third part concerns the units that have to be derived from the data from the first two parts. The main statistical units to be represented are those defined in the Regulation on Statistical Units. So far these are the only ones that have been modelled. However, the definitions in the Regulation are not always specific enough for the purposes of the model, so sometimes choices have to be made as to the interpretation of the units. Alternatively, different interpretations of a unit may be translated into different algorithms for deriving the unit. In addition, an analysis of the user needs may reveal the need for new types of units supplementing the set of units defined in the Regulation.

9. The comprehensive model is shown in Figure 2 as an ER (entity-relationship) model[1]. The three parts are linked at the top by the Enterprise Group (EG), which delineates the universe[2]. The entities defined for collection purposes are the organisational unit (OU), the OU's local parts (LOUs), the OU's activity parts (KOUs), the cross-section of the LOU and the KOU (LKOU), and the accounting unit (AU). Other entities do not need explaining, except for the so-called Building Block (BB), which links the three parts at the bottom of the figure. This entity is the smallest one of the model and is the cross-section of the smallest potential collection entity, the LKOU, and the legal unit. It is homogeneous in respect of activity, of organisation (i.e. belongs to only one OU), of location, of bookkeeping system, and of legal organisation (i.e. belongs to only one legal unit). As a consequence, all statistical units can be expressed as a combination of BBs.

[Figure 2]

10. The attributes of the entities are not shown in Figure 2, but they include a wide range of items, needed for the derivation of statistical units as well as their activity codes. Important attributes for the OU are variables indicating its degree of autonomy in respect of decisions concerning the production process, and variables describing its economic activity at a detailed level, such as inputs, outputs, type

1

of process, origin of the inputs and destination of the outputs. The attributes of the AU specify what types of variables are recorded in its bookkeeping system. Differences in variables recorded result in different AUs.

11. The ER model of the simplified version is fairly similar to the comprehensive model, the main difference being the level of detail of the attributes. In particular the attributes specifying the degree of autonomy of the OU and those specifying the bookkeeping variables have been compressed.

THE USE OF THE MODEL FOR STATISTICAL UNITS

12. If data are collected in accordance with the model and supplemented by administrative data on the legal unit structure, the model can be used for the derivation of statistical units. For this purpose, a set of algorithms has been developed. The algorithms take as inputs the collected and administrative data as defined in the model, and have specific statistical units as output. The inputs comprise the entities, their attributes, and the links between the entities. However, it should be noted that in practice not all information defined in the model may be available, not even for the simplified model. For instance, a breakdown of the OU by activity may not always be feasible.

13. The first algorithm to be applied is the algorithm for the enterprise group, because it defines the universe. This algorithm makes use of information on ownership relations between legal units as an approximation for control. The other algorithms have to be applied in a certain order, because some types of statistical units are defined in terms of other statistical units. The KAU, for example, is defined in terms of the enterprise.

14. As an illustration, the algorithm of the enterprise works roughly as follows. The enterprise is defined in the Regulation as “the smallest combination of legal units that is an organisational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources”. The algorithm starts with a top-down examination of the organisation structure of an enterprise group, establishing what are the smallest OUs that are autonomous in the sense of the enterprise definition. This in turn is derived from the list of autonomy attributes of the OUs. When the appropriate level has been found, the relationship with legal units is examined in order to check that the OU consists of one or more (whole) legal units. If not, the next higher level is checked, and so on. The algorithm includes a provision for situations where ancillary activities are shared by several enterprises.

15. The work on algorithms so far has shown that a fair amount of interpretation of the definitions of statistical units is required. For instance, the expression “a certain degree of autonomy in decision-making” in the definition of the enterprise lacks precision. But assumptions are also inherent to the model itself. One of the assumptions put into the model, for example, is the requirement that OUs form a strict hierarchy.

THE USE OF THE MODEL FOR CLASSIFICATIONS

16. In the previous chapters we explained how businesses can be described in terms of the model and statistical units derived. Using the information included in the model description of these objects, we can classify them according to existing activity classification systems such as NACE Rev.1 (and national extensions). In addition, we can also use the model to design and maintain alternative activity classifications. However, the model itself does not prescribe the choice of classifications; the systems to be used and built ultimately depend on the user needs.

Classification of a statistical unit according to NACE Rev.1

17. Once a statistical unit is described in terms of the model, i.e. a combination of BBs including their attribute values, we can automatically classify this unit according to a given activity classification system. To illustrate this, we briefly describe the algorithm to derive the NACE Rev.1 code of a given statistical unit.

18. First, combinations of BBs belonging to the unit are formed in order to describe overall processes instead of separate activities at the most elementary level (see also [3]). Starting with the market-oriented BBs of the unit the product flows between BBs are traced (using the known destination and origin of the product flow of BBs). Then the values of the attributes of BB-combinations, such as input products, process and output products, are determined. Using a look-up table with all possible combinations of attribute values, designed and maintained for this purpose, the corresponding NACE code can be obtained. In order to bundle similar processes, BB-combinations with the same NACE code are put together. As the NACE classification includes a few "combination classes"[3], these cases are identified, resulting in further bundling where appropriate. Finally, in order to determine the principal activity of the statistical unit we follow the so-called top down method as laid down in the Introduction to NACE Rev.1. The information needed to determine (an approximation of) the value added of activities is included in the model description of the BBs, i.e. employment and turnover.

Design and maintenance of classification systems

19. There are various criteria to delineate and order classes of activity classifications, such as output products, input products and processes. The needs of users with respect to these criteria are diverse and can even be conflicting. Therefore, different activity classifications may be required, each of them to be expressed in terms of the model. However, considering the need for coherence and standardisation, a set of mutually *related* activity classifications might be a solution. This has to be done in such a way that individual classes do not get too small, since this would lead to problems with mass and/or disclosure.

20. There are various methods to design hierarchical structures, for instance, by means of cluster analysis or an association criterion. An important aspect of the obtained classification systems is the heterogeneity within and "distances" between classes. Measurement standards can be developed and

expressed in terms of the model. This implies that the model can be used in the measurement of the quality of different classification systems.

CONCLUSIONS

21. In the preceding chapters three main uses of the model were described:

- Use as the basis for defining statistical units.
- Use as the basis for defining classes, classification rules and homogeneity measures.
- Use as a tool for designing classification systems and updating them.

22. These uses may be the most obvious ones, but much more is conceivable:

- Use as a tool for quality management. The measurement of the quality of classifications, both in application and design, is notoriously difficult, and the same is true in respect of statistical units. The model makes it possible to define precise quality measures.
- Use for comparison and harmonisation of statistical units and classifications. If different countries express their definitions and practices in terms of the model, the differences can be identified and possibly quantified. If harmonised definitions of units and classes are specified in terms of the model, the European statistical standards would gain in precision, thereby fostering harmonisation.
- Use as a framework for integrating administrative data in statistical data collection systems. At a national level the model, or a national variant, can be used to compare and harmonise the data that are collected by different agencies, possibly resulting in national data collection strategies.

23. It is clear that the model has a high potential, but implementing the model will not be easy. In the second half of 2001 the model will be tested on the basis of data collected by means of the DPQ. The data collection will be very limited, both in respect of quantity and types of businesses and activities covered. However, the general feasibility of the data collection and the applicability of the algorithms for the statistical units and coding will be clear by the end of the year. If the results are positive, it will be up to Eurostat and the NSIs to realise the full potential of the approach in the coming years.

References

[1] Knight, D., CLAMOUR objectives and achievements, paper prepared for the ETK-NTTS 2001 conference, Crete, Greece, 18-22 June 2001.

[2] Council Regulation (EEC) N° 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community.

[3] Bijwaard, H., et al, A model to describe the activities of a business, research paper 0040, Statistics Netherlands (CBS), 2000.

NOTES

[1] A UML version is currently being developed.

[2] This unit is also listed in the Regulation on Statistical Units. In the model it plays a dual role.

[3] Examples are (a) mixed farming (01.3) next to growing of crops (01.1) and farming of animals (01.2) and (b) hotels with restaurants (55.11) next to hotels without restaurants (55.12) and restaurants (55.30).
