Distr.
GENERAL

CES/SEM.50/3
15 April 2003

Original: ENGLISH

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (Eurostat)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE/Eurostat Seminar on Business Registers
(Luxembourg, 25-26 June 2003)

QUALITY IN THE SWEDISH BUSINESS REGISTER

Invited paper submitted by Statistics Sweden*

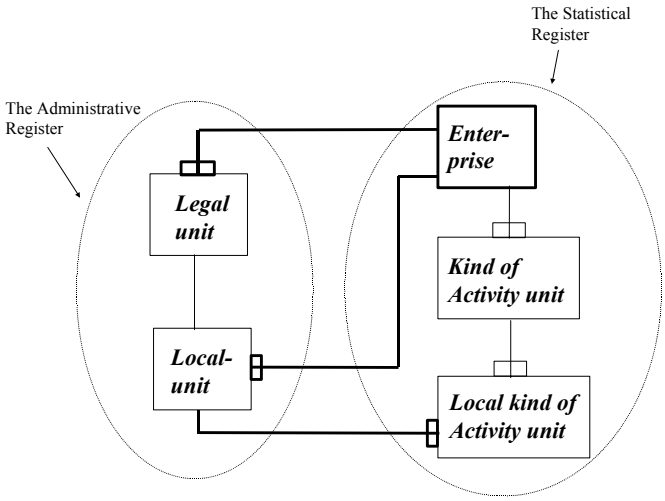
1. INTRODUCTION

1. The Swedish Business Register contains an administrative part or public part and a statistical part. The administrative part consists of legal units and local units and all register variables except some which are confidential (number of employees and turnover). The administrative part of the register with public variables is used for commission. The statistical part of the register consists of enterprise units, local units, kind of activity units and local kind of activity units. There is always consistency between the administrative part and the statistical part of the Swedish Business Register.

* Prepared by Mr. Henrik Romanov and Joakim Lundin, Statistics Sweden.

Figure

Units and parts of the Swedish Business Register, FDB



2. The quality of the Swedish Business Register is to a great extent dependent on the sources available. In Sweden, the Tax Authority is a very reliable administrative source and they update the register once a week. They update all legal units including information regarding registrations in the VAT-register and Payroll tax-register. In addition, the Tax Authority updates several other variables in the Swedish Business Register's administrative part.

3. The fact that there is one very good source for the updating of legal units is not enough to ensure the desired quality level. Maintenance of the business register therefore involves the use of other sources like questionnaires, information from companies which record address changes and feedback from statistical surveys. Information from Swedish institutions which organize periodic visits to local units to check workplace environments are also used as a source. Currently, approximately ten different sources are used to update the Swedish Business Register. Apart from that, the register is also updated with information from our own manual inquiries directly in the application. Manual inquiry is especially important for delineation of the statistical units; enterprise unit, local unit, kind of activity unit and local kind of activity unit.

2. QUALITY AND QUALITY INDICATORS

2.1 The sources are the base for quality in the Swedish Business Register

4. The quality of the sources is a useful indicator of quality in the register. A thorough description of how the sources obtain and transmit new information can give a good idea of the extent to which some of the quality criteria described in the EU recommendation manual can be achieved. For example, an estimation of the quality criteria timeliness can be made, even for every variable, provided that knowledge exists about which sources update which

variables and how often the sources updates the business register. Another example is that information about the sources can be useful in estimating over and under coverage. If information is available on how long it takes for the source to update the registration of new units, and about the average number of updates during a time period, this could give a measurement of coverage due to the time lag. A more thorough explanation of these indicators is presented under the headline "Suggested quality indicators used in the Swedish Business Register".

2.2 To find good quality levels

5. It would be ignorant to assess whether quality in a business register is good or bad without establishing specifications. To be able to evaluate quality in the register, the level of user demands has to be set and specifications established on what part of the register this demand counts for. For example, a measurement may indicate that activity codes are missing for about 20 percent of the enterprises, which can appear to be a great quality deficiency. But a closer examination might show that the units lacking the activity codes have only one or zero employees which makes them less important for most statistical surveys. This shows the importance of setting up acceptable quality levels and also the importance of doing this in cooperation with the users.

6. In the end quality objectives also have to be in accordance with the resources available. When acceptable quality levels are established, different ways to assess the achievement of the objective must be found. But how can quality really be measured for something as complex as a business register? One way of doing this is to use quality indicators.

2.3 Suggested quality indicators used in the Swedish Business Register

7. A helpful way to get a sense of the quality in a business register is to use the small pieces of information that are available. You can compare the process with doing a jigsaw puzzle. The first task is to gather pieces of quality information. Then you have to put them together to be able to get the picture of the total quality in the business register. The picture will be introduced to the users who determined the quality levels. The pieces are called quality indicators because they indicate quality for a certain part of the register at a certain time. The final picture created from the pieces is called the quality declaration.

8. The difference from a real jigsaw puzzle is that the pieces here have to be produced by the maintenance staff and that there might not be enough money or the possibility to create all of the pieces. Some of the questions that need to be answered are: When is the picture clear enough? Do we need more, better and more costly indicators or is this sufficient to satisfy the users?

9. Some quality indicators give a very strong indication of the quality, while others give only a tiny bit more knowledge than what we had before. Some indicators might not say anything about quality by themselves; however, combined with other indicators, they can provide a much closer indication of the true quality of the business register. Below, a number

of quality indicators are introduced. Some of them are used at Statistics Sweden today. Some are planned to be used in the future.

2.3.1 Indicator no. 1: Counting of units (See annex)

10. The flow of information into the business register should be checked frequently. This is for both units and variables. One way of doing this is to count all units at different points in time. If the number of units have increased or decreased in an abnormal way this would lead the maintenance staff to examine the reason why. To decide what is an abnormal increase or decrease, the number of units has to be examined together with previous values. Another important purpose for counting units is to give the users information about the range of the register. But it is also important to know the number of units to be able to calculate another quality indicator called coverage (Se chapter "Coverage").

11. To indicate quality in the business register, all units should be counted at least once a year. In Sweden the count is made four times a year. It is especially important that the enterprise unit, local unit, kind of activity unit and local kind of activity unit be counted. In addition, the Swedish Business Register has also decided to count the number of enterprises with several locations, enterprises with one local unit, complex enterprises, local units divided into local kind of activity units and the number of enterprises divided into kind of activity units.

12. A big change in the number of units could affect the usage of the register. Users may have to remake their survey or be forced to change their sample size. The reason for this abnormal change in number of units could either depend on the source, for example a change in legislation affecting the tax authority source, or due to some technical error in the updating procedure.

2.3.2 Indicator no. 2: Number of updates (See annex)

13. Another way of controlling the flow of information to the business register is to count the number of updates made per variable and unit. If the number of updates for a certain variable during a fixed time period is extremely high or low the reason of this has to be examined. The number of updates must be examined during a number of time periods to be able to get a good estimate of what is a normal number of updates.

14. It is also recommended to link the number of updates to different sources. This is especially useful in the evaluation of sources. Some examples of this are:

- Changes in the maintenance could be followed up. When the maintenance staff instructed the statistical surveys to update their feedback directly in the register application, they were given a source code of their own. It was useful to be able to see the number of updates, which variables they updated, and when they updated the register.

- The number of updates could help the users to identify the sources which are updating the variables used in their survey.

- The source code and the number of updates help to determine the timeliness of the variables (see chapter “Timeliness”).

- The number of updates along with number of units could be used for the calculation of coverage (see chapter “coverage”).

15. The Swedish Business Register has decided it is worthwhile to count the number of updates regarding status of local units. Therefore a count is made of the number of:

- Activated local units and inactivated local units

16. The number of updates on the following variables of local units is also worth counting. At the Swedish Business Register a count is made of the number of updates for:

- Number of employees, location, address, activity code, phone number, fax number and e-mail.

17. The number of updates on the status of enterprise units is also counted at the Swedish Business Register. A count is made of the number of:

- Activated enterprises and inactivated enterprises,

18. The number of updates on the following enterprise variables is also worth counting. At the Swedish Business Register, the number of updates is counted for:

- Number of employees, address, activity code, phone number, fax number, e-mail and name.

19. If the business register is particularly dependent on one source it is interesting to count the changes of variables in connection to the updates from that particular source. This is important because the quality of the register is to a large extent dependent on the correct handling of the update of information from this source. In Sweden, the Tax Authority is a very important source; therefore, errors in the transmission of information from this source to the business register could seriously harm quality if they are not detected at an early stage.

20. In the same manner as for counting the number of units and the number of updates, it is necessary to determine intervals for normal frequencies when updating the register with the important source. It is also important to investigate the reason why, if so, the number of updates is lying outside the interval. The reason might be the same for this source as for any source. This may be, for example to name a few, a change in Swedish taxation law, another way of transmitting the new information or sometimes the reason may simply be a technical problem.

2.3.3 Indicator no. 3: Timeliness (See annex)

21. Previous chapters described how it could be possible to provide good quality information by counting the number of updates and to link this to the source. Provided it is

clear which sources are updating which variables and the manner by which the sources capture information, it is possible to obtain information about the timeliness of the variables. In some cases the source captures the new information at the same time that this occurs in reality. But sometimes there is a time lag between reality and the moment of capture.

22. In Sweden, timeliness is introduced for the most important variables for enterprise units and local units. The indicators of timeliness drawn from the Swedish Business Register are to a great extent dependent on our knowledge of how the important sources (Tax Authority and our own questionnaires) capture new information and update the register.

2.3.4 Indicator no. 4: Coverage (See annex)

23. In calculating the coverage of units in the business register, the number of updates (introduced in an earlier chapter) could be useful. The Swedish Business Register almost exclusively uses the source Tax Authority for updating the status of enterprises. This knowledge and the knowledge of how often the Tax Authority updates the register, together with the number of activated units and inactivated units per update and the total number of units, is enough to calculate measurements of coverage due to the time lag.

24. To obtain a measurement of over and under coverage of enterprise units the first thing to do is to calculate the average number of activated and inactivated units per week using the indicator number of updates. The reason that the average is calculated per week is that the Tax Authority updates the register once a week. The calculated average is then divided by the number of units. The result is a measure of over and under coverage due to the time lag in the updating procedure. A measure of coverage could be calculated for other units in a similar manner. In Sweden, the enterprise unit is most suited for this kind of measure. Since the other units are not so dependent on the Tax Authority for status updates, this could make the calculations of time lags for these less reliable.

25. Except for the time lag factor, over and under coverage is also related to shortcomings in the sources. One example is that the Tax Authority fails to register enterprises active in the black market. Another example is that the questionnaires, feedback and other sources may fail to notice all local units. A third example is that the Tax Authority is slow in deregistration of VAT, which affects over coverage. To get closer to the true value of over and under coverage, these source shortages must be aggregated and added to the average value of activated and inactivated units.

26. In Sweden, over and under coverage for enterprises and local units are calculated four times a year along with a comparison with earlier measurements of coverage.

2.3.5 Indicator no. 5: Completeness (See annex)

27. In this paper, completeness refers to the coverage of a variable for a certain kind of unit, for example, the percentage of local units in the register with information regarding activity code. In Sweden, the register system requires a value for some of the variables, which means that these variables can never be blank or missing. These variables, for example employees and postal address, are 100% complete. Activity code, visiting address, phone

number, fax number, and e-mail are examples of variables where the completeness is far from 100%. Besides activity code, the completeness might not be that important for the statistics. But the administrative part of the Swedish Business Register from which information about enterprises and local units is sold, is on the other hand very interested in completeness for every variable. It might be valuable to have the ability to inform the customer that 10% of the enterprises with less than 5 employees lack phone numbers or that an e-mail address is only available for 3 % of the local units. It is important to notice here that for some variables it is impossible to reach 100% completeness because some units, for example, do not have a valid visiting address at the place they work (mostly at the countryside) or because some units have a secret phone number.

28. In Sweden completeness is calculated four times a year and compiled in a table with some of the most important variables for every unit. The completeness statistic is divided into different size classes, for example, by employees or turnover.

2.3.6 Indicator no. 6: Accuracy

29. When accuracy is discussed in this chapter it is with regard to variables. When completeness is calculated for a certain variable, it is not likely that all existing variable values are accurate. To find out whether the values are accurate or not, they have to be examined. The ability to measure the accuracy of the content in a business register is dependent on the amount of resources available. If there were unlimited resources, all units could be examined by the best register staff over and over again. But if resources are scanty, which is most likely, one possibility could be to estimate the accuracy by examining only a sample taken from the register. The sample would preferably be controlled with the enterprise via phone. These random controls of accuracy would probably also involve much time and monetary cost; therefore, more cost-effective ways of measuring the accuracy has to be found (More about this is described in chapter 4). At Statistics Sweden, no measurement of accuracy is available at the moment.

3. COOPERATION WITH INTERNAL AND EXTERNAL USERS

3.1 Quality information from statistical surveys

30. As mentioned before, it can be both costly and time demanding to obtain a measurement of accuracy for the variables in the business register. One way of reducing cost and time is to use statistical surveys that already exist. These surveys could work in combination to collect statistical information and to check the accuracy of register variables.

31. A precondition for being able to use a statistical survey as an indicator of the accuracy in a business register is that the survey be based on a sample or a total examination of the business register. If it is not, the units and variables might not match that of the business register and the definitions of variables may differ as well.

32. If the survey is based on the register, the units will have the same definition in the survey and in the register. The likelihood that there will be similar definition of variables also

increases, but this is not certain. For example, the postal address used in the statistical survey could be directed to the contact person. This person may be located at a different address from the address indicated by the sampled unit which in fact is the address that the business register is interested in. Another example of different definitions of variables is the number of employees, which often differ between the survey and the business register. This is often a fact, at least for Statistics Sweden.

33. When questionnaires are to be sent to the sample group as part of a statistical survey, it is necessary to think about using the survey to gather information regarding accuracy in the business register. One way of doing this is to count the number of incorrect variable values for variables already printed in the questionnaire, for example postal address. In this case, the percentage of postal returns can be calculated, which could give an indication of the accuracy for the variable postal address in the part of the business register from which the statistical survey drew its sample. The accuracy for the activity code could be measured in a similar manner. Many questionnaires explain what kind of activity the statistical survey is examining, which often triggers the unit being surveyed to react if they think that this activity does not correspond to what they actually do. In this procedure there is no spoken request for the examined unit to react when a register variable is wrong, which makes this alternative to measuring accuracy a bit vague. The good thing, however, is that it is relatively easy and does not cost much to implement.

34. Another way of using the statistical survey is to produce a separate paper with register variables. This paper is enclosed with the questionnaires and the examined unit is asked to change wrong variable values and to verify the correct ones. One advantage with this procedure is that the examined units are urged to change incorrect information. Another advantage is that there is the possibility to add variables that the statistical survey does not currently use. The disadvantages are that the person responsible for the survey may perceive the enclosed paper with register variables as distracting focus from the survey and makes it more troublesome for respondents to answer the questionnaire.

35. The statistical surveys are often conducted on samples drawn from the register, which does not necessarily mean that the result has to be accounted for in all of the units in the register. Most likely the sample is stratified by some variables, for example activity code or number of employees. This means that the result of the accuracy measurement can be accounted for per size class or per group of activity codes. If it is possible to examine accuracy with the help of only one survey, the result will be rather narrow. That is why it could be very interesting to use many different surveys to gather information of accuracy in different groups of activity codes and in different size classes.

3.2 Other sources for measuring the accuracy in the business register

36. If the business register has an integrated statistical and administrative part (as it does in Sweden), it is also possible to use the commission of the register for an accuracy check of the variables. Suppose that those maintaining the business register are selling addresses or phone numbers from a certain strata. The maintenance staff then could give the customer a reason to give feedback concerning the number of correct or incorrect addresses and phone numbers. The incentive to do this could be, for example, an offer to the customer of a small

amount of money or discount on future purchases. This way of calculating accuracy contains a time lag. When analysing the results, consideration has to be given to the amount of time between the purchase from the register to the customer's contact with units within the register.

37. Another way of getting a better idea of accuracy in the business register is to use existing knowledge about the sources. For example, when an enterprise becomes active in Sweden, the Tax Authority always decides the activity code. So, the quality of the activity-code is to a great deal dependent on the coding skill of the Tax Authority. Therefore those maintaining the business register could draw a sample from the new units and examine how many codes are correct and how many codes are incorrect, just to get an idea of the accuracy of the activity codes. To make it even more cost-effective, the maintenance staffs could examine only those units that already have to be examined (the maintenance staffs always investigates activated units with more than 10 employees). In this way every source could be examined and evaluated in order to determine their usefulness and the accuracy of the variables they update.

4.3 Other aspects of user influence on quality in the Swedish Business Register

38. In Sweden we have started a Business Register Council. The Council consists of users of the Swedish Business Register within Statistics Sweden. The Council discusses larger issues concerning the content of the register. The chairman of the Business Register Council, also the key decision maker, is the head of the Economic Statistics Department in Statistic Sweden. Today the Council consists only of internal users, but it would also be good if external users could be gathered in a similar council to get some idea of their views on the quality of the business register.

39. The work in creating statistical units is sometimes hard and time demanding. To get a good statistical view of the delineation of statistical units, Statistic Sweden have created a group comprised of participants from almost every statistical program using the business register including some maintenance staff. This group gathers once or twice a month to go through delineations of new complex enterprises and changes in variables of existing complex enterprises. This is very important in order to prevent the gap from becoming too big between the maintenance staffs and the statistical programs understanding of the delineation and definition of statistical units.

Annex

Indicator no.1 - Counting of units

Explanation: SAMU stands for our sample moments

Unit	Mars-Samu 2002	Nov-Samu 2001	+- number	**+- %**
Total number Legal units	1 567 537	1 544 320	23 217	1,5
Number of LU with several locations	8 009	8 175	-166	-2,0
Number of local units	909 009	908 976	33	0,0
Number of local units divided in LKAU	41	33	8	24
Number of enterprises	830 158	828 740	1 418	0,2
Number of complex enterprises	47	47	0	0
Number of KAU-divided enterprises	72	73	-1	-1,4
Number of LKAU-divided KAU	7 995	8 159	155	1,9

Number of enterprises in different size classes

Sizeclass	Antal FE
1-9 employees	179 821
10-499 employees	33 275
> 500 employees	855
0 employees but > 50 SEK millions turnover.	536

Indicator no. 2 – Number of updates

Number of updates for legal units

Period 2001	3/1-21/11	Total/11	Total /46
	Total	Per month	Per week
New reg. LeU	72 395	6 033	1 392
De. Reg. LeU	69 477	5 789	1 336
Difference	2 918		
Activated Ent.	66 447	5 537	1 277
Inactivated Ent.	53 150	4 429	1 022
Difference	13 297		
Number of emp.	225 959	18 830	4 345
e-mail	3 717	309	71
Fax	7 889	657	151
Name	128 087	10 673	2 463
Activity	192 232	16 019	3 696
Postal address	486 873	40 572	9 362
Phone	114 497	9 541	2 202

Number of updates for local units

Period	3/1-21/11	Total/11	Total/46
	Total	Per month	Per Week
New reg.	75 740	6 311	1 456
De. Reg.	64 856	5 404	1 247
Difference	10 884		
Number of emp.	168 561	14 046	3 241
location	213 657	17 804	4 108
nickname	8 408	700	161
e-mail	5 689	474	109
Fax	15 678	1 306	301
activity	75 000	6 250	1 442
Postal address	218 044	18 170	4 193
Phone	50 122	4 177	963

Indicator no. 3 - Timeliness

Legal u./Enterprise			Local unit		
Variables	Timeliness		Variables	Timeliness	
Name	1-3 w		Postal address	1-3 w	1-lu
Postal address	1-3 w			1-12 m	s-lu
			location	1-3 w	1-lu
Area code	1-3 w			1-12 m	s-lu
			Area code	1-3 w	1-lu
Number of employees	5-16 m	1-lu		1-12 m	s-lu
	1-12 m	s-lu	Number of emp.	5-16 m	1-lu
Activity code	1-13 m			1-12 m	s-lu
Phone	1-3 w		Activity code	1-13 m	
Turnover	11-22 m		Phone	1-3 w	1-lu
Real estate	11-22 m		Nickname	1-12 m	s-lu

S = several
Lu = local units
M = months
w = weeks

Most of the variables in the table are updated by the Tax Authorities, the address changing company or by our own questionnaires. Name, turnover and real estate class are updated in batch and turnover and real estate class are updated only once a year.

Indicator no. 4 – coverage

Counting of over and under coverage for enterprise/legal units due to time lag.

The total amount of legal units in the register is 829 449 units. Every week 1051 units are added to the register. We update the register with new units every week so we can say that we have an under coverage of 1051 units and approximately 0.12 % of al units.

At the same time 1296 units or 0.15 % of the units are activated describing the over coverage.

Note that the figures are from May 2001.

Indicator no.5 - Completeness

Variable	Active LeU	Active LeU in %	Variable	Active LU	Active LU in %
Name	830 805	100	Postal address	909 009	100
Postal address	830 805	100	Location	869015	95,6
INST-code	830 805	100	Area code	909 009	100
Number of em	830 805	100	Number of emp	909 009	100
Activity	744 872	89,7	Activity	823 075	90,5
Phone	531 827	64	Phone	609 474	67,0
Fax	13 210	1,6	Fax	39 097	4,3
E-mail	5 825	0,7	E-mail	11 317	1,2
Area code	830 805	100			
