Statistics Sweden

Statistiska centralbyrån

The future development of the Swedish register system
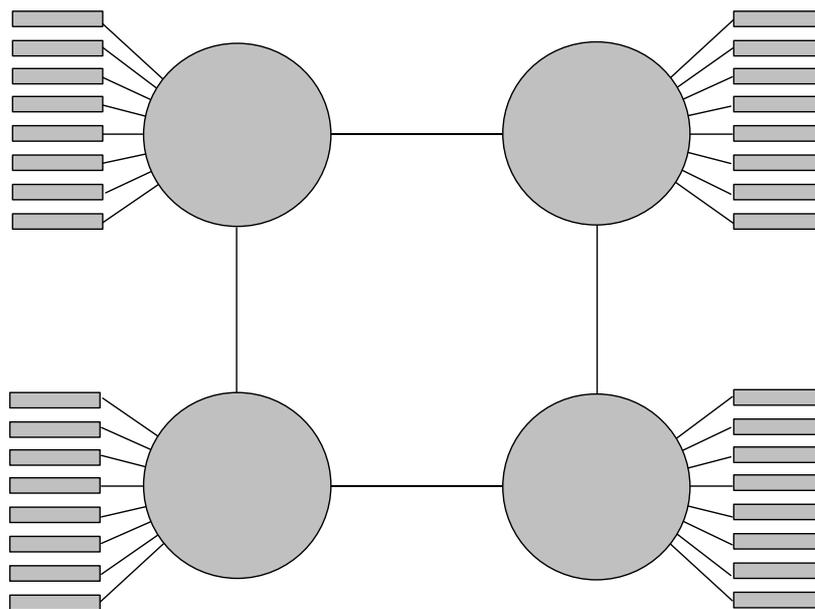
Final report of the Register Project

and decision of the Director-General

R&D Report 2001:1

Research – Methods – Development

The future development of the Swedish register system

_____

Contents

Preface

Swedish official statistics have long been based upon the extensive use of administrative data in the construction of statistical registers. These registers are frequently used both as the basis of our sample surveys and for publishing different types of register statistics.

Despite this register tradition, however, Statistics Sweden has not commonly thought in terms of a statistical register system – the registers have been treated independently. The task of the Register Project has been to elucidate register-statistical methodological work, to develop clearer thinking in terms of a system and to bring about greater co-operation between the different parts of the system. The setting up of the project was an indication that the register system was to be regarded as a strategic resource.

The project group, of which this constitutes the final report, has consisted of Elisabet Andersson, Evert Blom (project director), Linda Ryen and Anders and Britt Wallgren. During a short initial period, up to April 1997, the project was directed by Staffan Wahlström.

In addition there has been attached to the project a contact group consisting of some 20 persons associated with different Statistics Sweden registers. This group has been of the greatest importance – we have been able to discuss register problems with persons who have years of experience of register statistics, and been able to spread knowledge of the system and of a number of registers of general interest through having members of the group and guest speakers tell of their registers and share their experience.

From the project there is to come a book of register-statistical work, which can be used on courses.

The report has been circulated for comment within Statistics Sweden and has been discussed by the directorate. This has led to certain corrections. The Director-General's decision regarding the report and its handling is to be found on the following page.

Decision of the Director-General 7 May 2001 with reference to the final report of the Register Project

Background

The final report of the Register Project was considered by the directorate on 13 March 2001 after it had been circulated for comment within Statistics Sweden. The Director-General declared that the groundwork had been well carried out and that the proposals were supported, subject to certain modifications, by the directorate. The modifications mainly concerned the proposal concerning general guidelines, which also should be reduced in number. The Director-General declared further that the handbook needs to be completed by the end of this year, also that the R&D Department shall be responsible for the future handling of the overall register questions.

Decision

Against the background of the internal circulation for comment and the discussion by the directorate it is decided:

- that general guidelines in accordance with Section 1 of the final report shall be valid

- that the register model in accordance with the conceptual model in Section 3 of the report shall constitute the basis for the description and further development of the register system

- that the departments and programs shall be responsible for the development and improvement of the register system on the basis of the proposals in Section 5 of the report; whereby each department shall have at least one methodologist specialising in register questions

- that a central register function shall be established within the Methods Unit of the R&D Department; whereby the department shall be responsible for the specific form of this function

- that the question of measures in accordance with the proposals concerning increased protection of register data shall be referred to executive management (staff)

- that the R&D Department through the Methods Unit and the IT Unit, in cooperation with other departments and units, shall attend to the proposals concerning further development; whereby the department shall see that these matters are dealt with in the normal process for the establishment of future plans regarding methods and IT

- that the handbook concerning register questions shall be completed by the end of 2001

- that the R&D Department shall be responsible for the further handling of the overall register questions, including the implementation and following up of present decisions

1. Summary

In this section the project's proposals are summarised. The decision proposals are grouped in accordance with the units, etc. responsible for their realisation.

Guidelines for statistical registers at Statistics Sweden

1. Statistics Sweden shall increase its use of administrative data so as to lighten the respondent burden and to be able to produce new statistics without an increase in this burden.

2. In all surveys there shall be consideration as to whether and what register data can be used. Register data can be used instead of asking, for quality control and as auxiliary information to reduce errors of sampling and non-response in sample surveys.

3. Statistics Sweden shall co-operate with authorities which submit administrative registers and if necessary attend to the need for statistical variables or for the adjustment of administrative data in order to improve the statistical value of the information. The contacts shall be co-ordinated when several departments and programs are involved. Quality information shall be passed back to the authorities in question.

4. Administrative data received shall be processed in a uniform way. Just as sample surveys are carried out in accordance with an accepted methodology, so shall Statistics Sweden's registers based on administrative data be established in accordance with accepted principles. This means that there must be regular co-operation across department and program borders concerning questions of method with regard to register statistics.

5. All statistics shall be quality-certified and well documented, as also shall the registers from which they are taken. Registers, which are a part of the register system, shall be documented in Metadok and there shall be a complete SCBDOK documentation.

6. Definitions of objects, object sets and variables shall be co-ordinated. There shall be standardised variables which have particular names, are defined and are easily accessible. They shall not be used with a divergent definition.

7. Every statistical register shall contain the reference variables required for linking the objects to the particular basic register.

8. The registers comprise the major part of Statistics Sweden's extensive microdata store. The latter shall be managed in such a way as to ensure that the secrecy and protection of privacy requirements are fulfilled.

9. If there is need to establish a permanent integration register a working party shall be set up with the task of ascertaining whether a new register is the best solution and of elucidating the areas of responsibility, rights and obligations of the programs involved.

10. Statistics Sweden's register system is a part of the statistical system and shall be usable by statistical authorities and others instead of building up and maintaining extensive databases with register information of their own.

Proposals

Central register function

A central register function with responsibility for questions of development and co-ordination shall be established in the R&D Department. To it shall be linked a network of staff from the rest of Statistics Sweden with experience in methods and production. The central register function shall be responsible for:

– overall questions of co-ordination and standardisation;

– the meta-data system and the development of documentation support;

– the establishment of a network for the exchange of experience and the development of new forms of register processing and competence regarding complex register analysis within the frame of the network.

Organisational questions with regard to this function shall be dealt with in connection with the R&D Department's investigation concerning a central register service function and the question of a reinforcement of the central standard function.

The IT Unit

The production and operational environment needs to be improved in order to underpin the development of the register system. It shall be the task of the IT Unit to:

– adopt an overall perspective and take firm steps to thin out the variegated tools for register processing and develop new general tools;

– make Statistics Sweden's data store more accessible and increase the capacity for handling large volumes of data;

– develop the meta-data system so that the documentation and use of such data are integrated with the production side;

– develop technical solutions making meta-data for standardised variables generally accessible and making it possible for such data to be introduced into production;

– arrange seminars concerning experience of different stages of production and registers.

The Methods Unit

Register-statistical methodological work shall be given the attention it so clearly deserves. It shall be the task of the Methods Unit to:

– develop methods for the editing of large registers;

– develop methods for the management of non-response with regard to register statistics;

– develop the methodology of record linkage;

– develop methods for the evaluation of objects and variables in the register system;

– implement a quality assurance system;

– devote particular attention to method problems regarding the basic registers.

The handling of the method questions shall be carried out within Statistics Sweden's general methods program.

Departments and programs

The putting into practice of the proposals for improving the register system shall take place on the department and program levels.

– Every department shall have responsibility for its basic register and for other registers within its field, together with special responsibility for a number of standardised variables.

– Every department shall be responsible for the keys between its statistical registers and the particular basic register.

– The part of the register system which has to do with statistics on persons shall be used as a model and source of inspiration for developing register statistics within the system as a whole.

– Each of the department methodology groups shall contain at least one methodologist responsible for register questions.

The proposals for the improvement of registers which are to be found in Section 5 shall be put into practice step by step but in good time before the 2005 population and housing census. This can be taken up in guideline documents addressed to departments and programs. Certain improvements call for co-operation between different departments and programs.

Secrecy and protection of privacy

Statistics Sweden's preservation of secrecy and protection of privacy within the register system needs to be reinforced.

– It should as soon as possible be investigated how persons indicated as requiring special protection of identity can be given increased protection within Statistics Sweden's register system.

– It should be investigated whether civic registration numbers can be replaced by serial numbers in one or more of Statistics Sweden's sensitive registers.

– The present rules for the giving out of microdata need to be made more precise. It must be brought to the attention of both users and staff that the same rules apply concerning corporate registration numbers as concerning civic registration numbers. The rules need to be made more concrete. Which variables are more sensitive and which are less so?

The Financial Unit

The Financial Unit should clarify and disseminate the rules for internal and external pricing of register services.

Launching

The project has already involved a great deal of work with regard to launching. Many seminars and discussion meetings have been held where the various ideas concerning register statistics have been presented.

Later there will come from the project a handbook concerning register statistics. Offering a description of the register-statistical system, it will be able to be used on both introductory and extension courses for staff working with register statistics, as well as being, it is hoped, a source of inspiration for persons working with registers.

The launching will continue with methodology seminars and courses.

2. Background, purpose and ways of working

Background

Practically all the surveys carried out by Statistics Sweden involve the use of one or more of Statistics Sweden's statistical registers. Developing the register system will enable us to perform new register-based surveys at low marginal cost and to improve our present surveys.

We are encountering new demands from the world around. The extensive social changes which are taking place increase the need for new statistics. Since, however, the non-response and the requirements regarding reduced response burden are increasing, at the same time as it is difficult to obtain larger appropriations, we cannot always meet these demands by means of new sample surveys. We must instead learn to make more efficient use of already existent register data. The 2005 population and housing census offers a good example of the new demands. Without traditional data collection we have to be able to carry out a wholly register-based census.

We estimate that registers account for about half of the economic value of Statistics Sweden's production. In spite of this, however, register-statistical production and methodological work have led rather an obscure existence. In contrast with what has been the case regarding sample-based survey statistics, which have decades of methodological development behind them, the need for systematics and well-thought-out processes in the case of register statistics has not received the attention it has deserved. In our opinion register statistics have a great development potential, and we believe that if our proposals are acted upon there is a good chance of increasing Statistics Sweden's productivity.

Register statistics and the development of the register system require cooperation across program and department borders. The old view of things whereby the different parts of the organisation have their own small register systems adapted to their own particular needs is an obstacle to thinking in terms of the whole – an obstacle to a unified register system for the whole of Statistics Sweden!

Purpose

The purpose behind the Register Project was the following:

"Statistics Sweden's different registers which are based on administrative sources shall be developed into a well-functioning register system. The various parts of the system shall be clearly defined, and through increased cooperation the system shall be made more efficient so that new integration registers can be established in order to meet new requirements on the part of the users. The register-statistical methodological work shall be rendered visible and be developed so that the quality of the register statistics can be described and improved."

The numbered and italicised conclusions and proposals which are to be found in the rectangular "windows" of this report represent our endeavour towards fulfilment of the said purpose.

Ways of working

The introductory register seminar was a great success and gave the project a fine start. In connection with the drawing up of various inventories we have been in close touch with register programs. We have also co-operated with other Statistics Sweden projects: the Database Project, the Louise Project, the Meta-data Project and the Platform Project.

A special contact group with some 20 members has functioned as the project's reference group, with broad representation from different registers. The majority of the registers were described at the meetings of the contact group, mostly in the form of seminar-type presentations. These meetings were greatly appreciated by the participants because of the information given about other registers – information furthering co-ordination and professional development. Furthermore the project has involved practical work with several registers for the purpose of elucidating various problems of method.

3. The register system – registers and keys

A statistical register system is composed of a number of different parts, and this section takes up registers and keys. Objects in different registers are linked by means of reference variables.

– The basic registers consist of object sets of fundamental importance. In the charts below these registers are indicated by circles.

– Other statistical registers contain statistically interesting variables.

– Keys between objects in different basic registers and between basic registers and other statistical registers are indicated by lines in the charts.
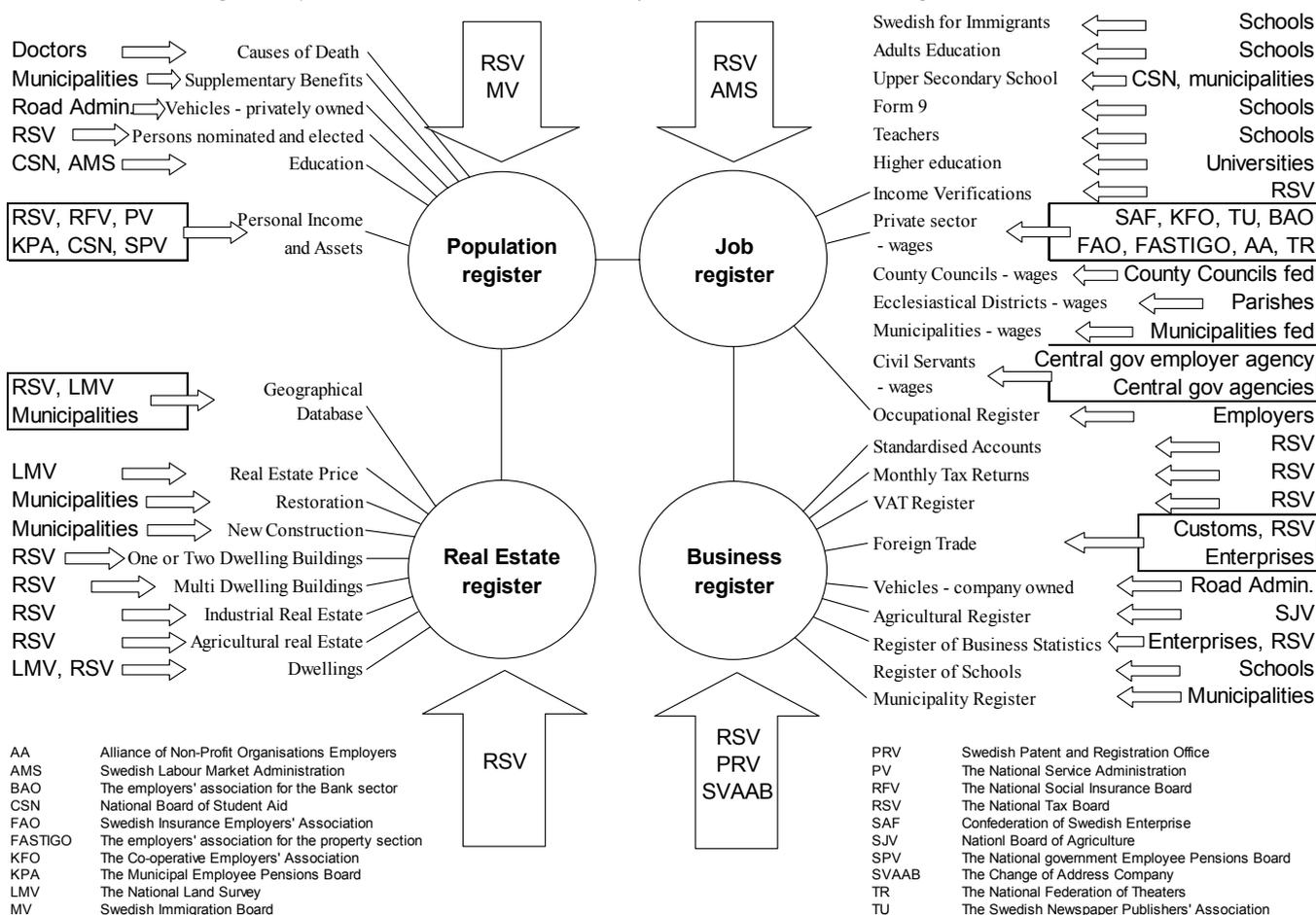
A complete register system also includes standardised variables, register-statistical methods, IT tools, meta-data and routines for the protection of privacy. These are discussed in Sections 4–9.

Within the Register Project we have been occupied in inventorying, structuring and attempting to understand the register system. Such concepts as "register system" and "register-statistical methodological work" represented a new way of thinking when the idea of the project was first presented in 1996. Since then we have developed this way of thinking, and we now describe the register system in the following way, presenting its construction step by step.

Step 1. Reception of data from different administrative sources

From authorities supplying registers different programs within Statistics Sweden (SCB) receive administrative data which is processed in such a way that a number of statistical registers can be established. These registers can then be linked to one another inasmuch as objects in different registers can be identified by means of the same reference variable. The following chart represents the input part of the system:

Chart 1. SCB's register system – how we receive data from authorities and organisations

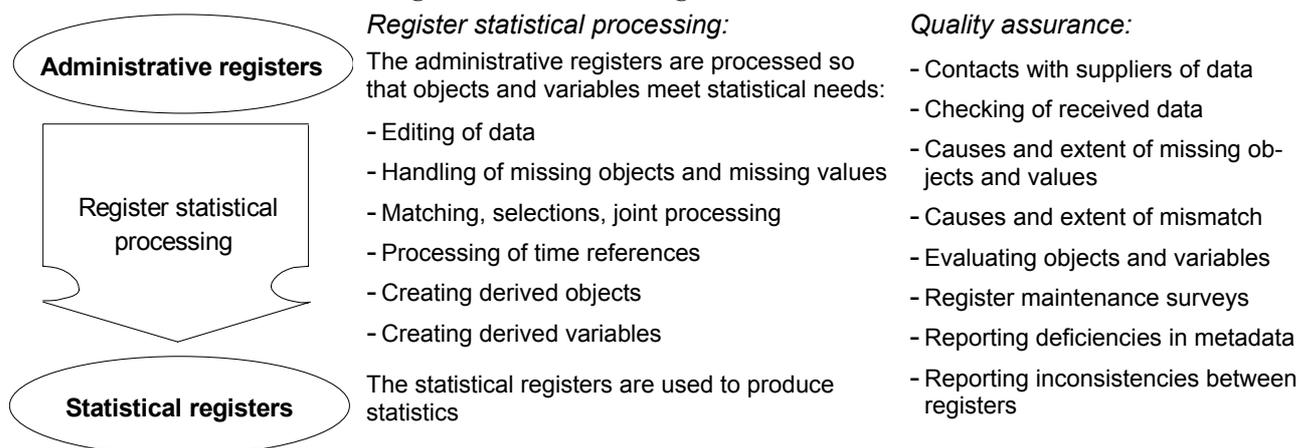|		
AA	Alliance of Non-Profit Organisations Employers	
AMS	Swedish Labour Market Administration	
BAO	The employers' association for the Bank sector	
CSN	National Board of Student Aid	
FAO	Swedish Insurance Employers' Association	
FASTIGO	The employers' association for the property section	
KFO	The Co-operative Employers' Association	
KPA	The Municipal Employee Pensions Board	
LMV	The National Land Survey	
MV	Swedish Immigration Board	
PRV	Swedish Patent and Registration Office
PV	The National Service Administration
RFV	The National Social Insurance Board
RSV	The National Tax Board
SAF	Confederation of Swedish Enterprise
SJV	Nationl Board of Agriculture
SPV	The National government Employee Pensions Board
SVAAB	The Change of Address Company
TR	The National Federation of Theaters
TU	The Swedish Newspaper Publishers' Association

It can be seen from Chart 1 that many programs have contacts with register-supplying organisations and that a large number of authorities and organisations are involved. Chart 1 warrants the following conclusion:

1. Statistics Sweden's supply of administrative data is organised and made secure through well-arranged contacts with register-supplying organisations. These contacts occur at different levels but should be co-ordinated within Statistics Sweden.

Step 2. From administrative registers to statistical registers

As a rule it is inappropriate to attempt to produce statistics direct from the administrative registers which have been received, since these registers are not adapted to the needs of statistics. Object sets, object definitions and variables need to be checked, and thereafter it may be necessary to do a certain amount of processing so that the register fulfils the requirements of statistics with regard to objects and variables. Such processing, the purpose of which is to transform one or more administrative registers into a statistical register, must be based on well-thought-out and standardised register-statistical methods. Chart 2 shows the various components of the methodological work.

Chart 2. From administrative registers to statistical registers

Administrative registers

Register statistical processing

Statistical registers

Register statistical processing:
The administrative registers are processed so that objects and variables meet statistical needs:

- Editing of data
- Handling of missing objects and missing values
- Matching, selections, joint processing
- Processing of time references
- Creating derived objects
- Creating derived variables

The statistical registers are used to produce statistics

Quality assurance:

- Contacts with suppliers of data
- Checking of received data
- Causes and extent of missing objects and values
- Causes and extent of mismatch
- Evaluating objects and variables
- Register maintenance surveys
- Reporting deficiencies in metadata
- Reporting inconsistencies between registers

These components are common to many of the register programs at Statistics Sweden. Therefore it is important that there should be exchange of experience and methodological development, this in order that well-established and well-documented methods be brought into general use in the way which is self-evident when it comes to Statistics Sweden's sample surveys. Chart 2 warrants the following conclusion:

2. The receiving of large volumes of administrative material and the editing of this material are important operations that many programs perform in the same way. These operations should be made more efficient by means of exchange of experience and the development of common tools and methods of editing. Similarly, methods and tools should be developed for non-response management, joint processing and evaluation. The responsibility for this should rest within the general methods program and the work should be carried out in close contact with the register programs.

Step 3. Establishment of further registers and uses

The registers under a) and b) in Chart 3 below have been established by use of administrative data from different register-supplying organisations. These registers form the basis for different statistical products.
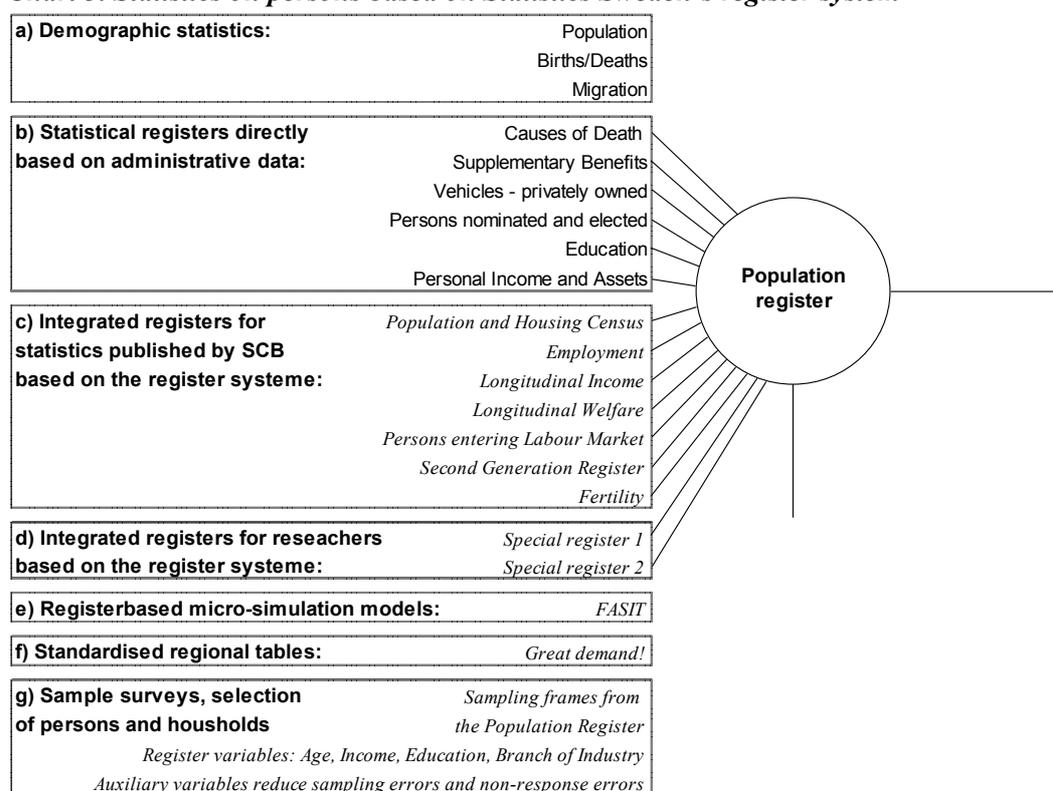
By jointly processing several registers within the system it is possible to create integration registers. These can be either temporary ones for specific tasks or permanent ones for new Statistics Sweden products.

The registers under c) and d) in Chart 3 are such integration registers, and they have been created without further data collection.

The capacity to perform joint processing is dependent upon how highly developed Statistics Sweden's register system is. The data store needs to be well-organised and there need to be efficient IT tools for different types of register processing. Our main proposal is that Statistics Sweden's register system be developed so as to facilitate the creation of integration registers, both longitudinal ones and others.

The part of the register system which concerns statistics on persons is that which most developed, and it can serve as a model for the other parts.

Chart 3. Statistics on persons based on Statistics Sweden's register system

a) Demographic statistics:	Population
	Births/Deaths
	Migration

b) Statistical registers directly based on administrative data:	Causes of Death
	Supplementary Benefits
	Vehicles - privately owned
	Persons nominated and elected
	Education
	Personal Income and Assets

c) Integrated registers for statistics published by SCB based on the register systeme:	*Population and Housing Census*
	Employment
	Longitudinal Income
	Longitudinal Welfare
	Persons entering Labour Market
	Second Generation Register
	Fertility

d) Integrated registers for reseachers based on the register systeme:	*Special register 1*
	Special register 2

e) Registerbased micro-simulation models:	*FASIT*

f) Standardised regional tables:	*Great demand!*

g) Sample surveys, selection of persons and housholds	*Sampling frames from the Population Register*
Register variables: Age, Income, Education, Branch of Industry	
Auxiliary variables reduce sampling errors and non-response errors	

Population register

This chart illustrates the possibilities offered by a system of statistical registers. The development of the register system as a whole will provide Statistics Sweden with a greater capacity for developing statistics based on integration registers. An improved register system makes it possible to improve Statistics Sweden's sample surveys in that questions can be replaced by register variables and the use of auxiliary information from registers can reduce errors with regard to sampling and non-response.

3a. The part of the register system which concerns statistics on persons should be used as a model and source of inspiration for the development of register statistics within the whole system. Demographic statistics for enterprises, real estate/buildings, employment, etc. should be developed. Simulation models and standardised regional tables should be developed within the whole system. The use of the registers shall be facilitated through making many register variables easily accessible for sample surveys.

3b. In addition to developing existent registers we must be constantly observant regarding new administrative sources which can be used for the creation of new statistical registers and new products.

Step 4. Statistics Sweden's register system – creation of the totality

Within the Register Project we developed the model of the system after having made an inventory of the existent registers. We studied especially which types of object were included in the different registers and which variables were used for linking objects in different registers with one another. We found that the registers could be divided into four groups in accordance with type of object. Within each group there is a basic register defining all the objects in the group.

The role of the basic registers is to define the objects in the system, whereby good object definitions and good coverage are decisive with regard to the quality of the system as a whole. The keys between the objects in different basic registers are particularly important when it comes to the possibility of creating integration registers with interesting content. The basic registers contain the statistically interesting object sets, whilst the statistically interesting variables are in the other registers. The variables which are important in a basic register are those which identify the objects and can be used to connect the objects to objects in other registers. Furthermore there is a need for time data regarding different events which affect the objects, so as to be able to create a stock of data with reference to a certain point in time or period. This is a prerequisite for demographic statistics. Chart 4 shows the registers which at present can be linked to their respective basic registers.

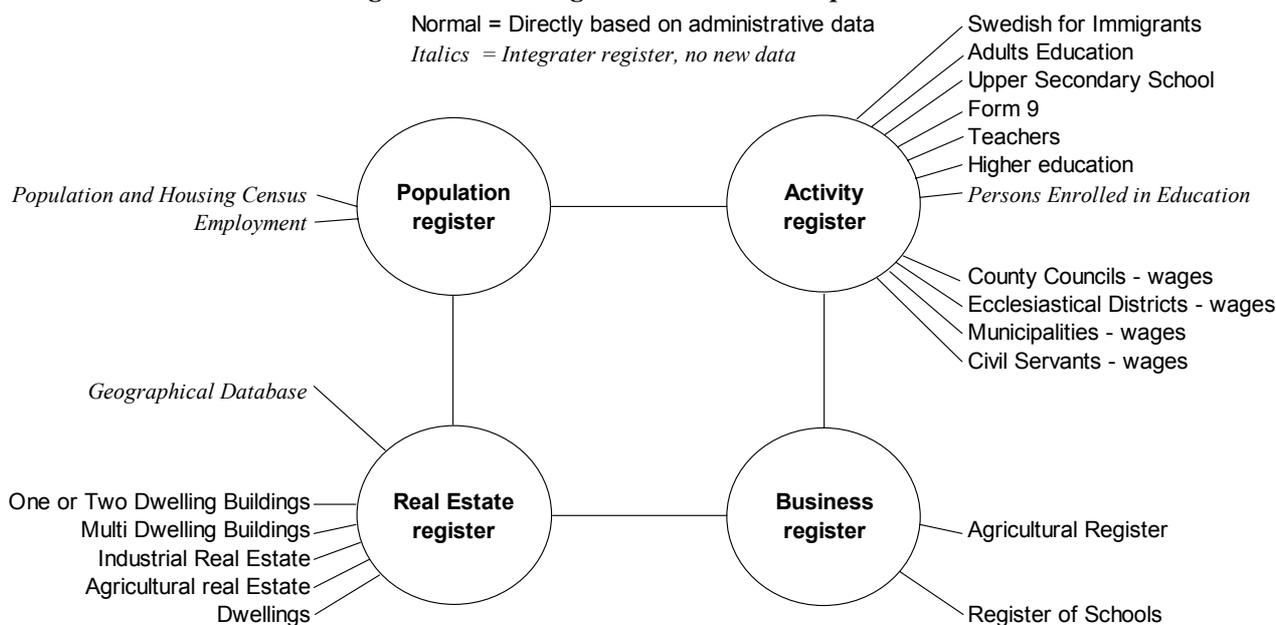*Chart 4. SCB:s present system, register by object types*
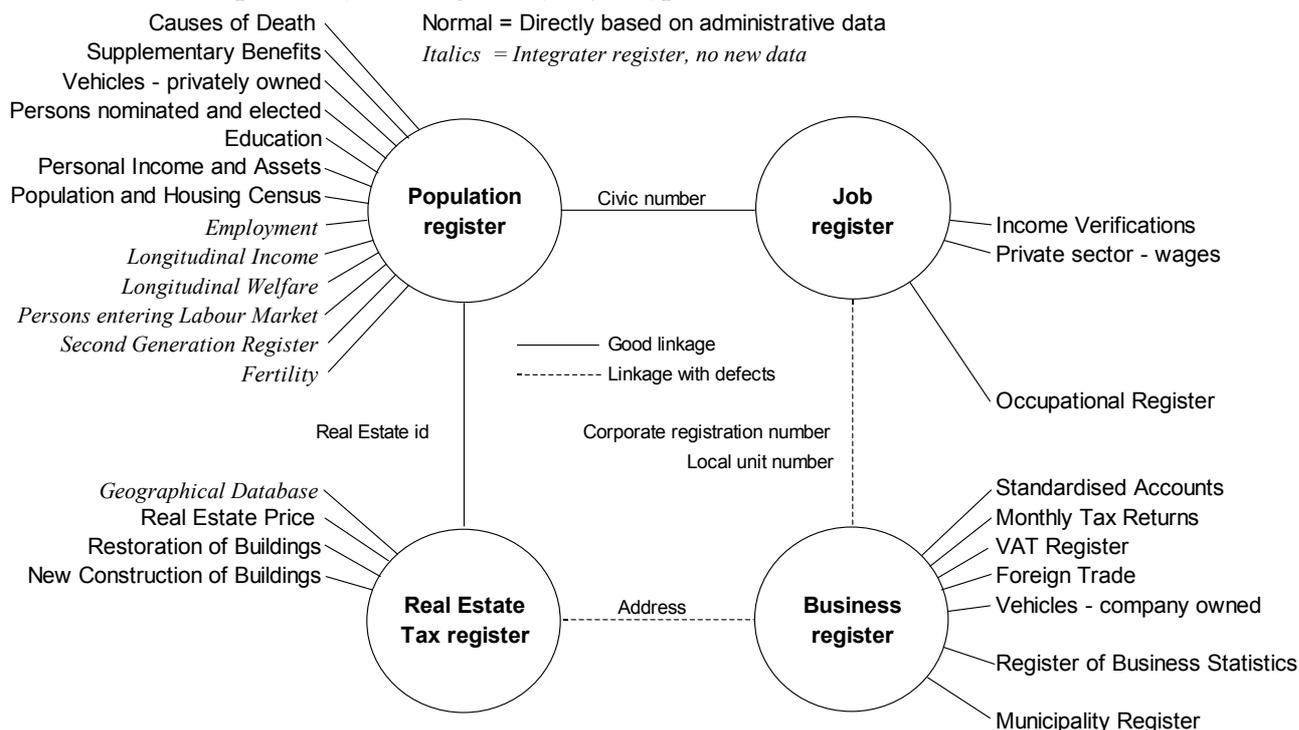
Of our present basic registers the Real Estate Register does not have the structure appropriate to an out-and-out basic register for objects within the area of real estate. In the first place it has so many variables as to make it unsuitably large for its role as basic register, in the second place there are coverage problems inasmuch as it is updated only once a year.

The Job Register should be made into a more general Activity Register able to function as a basic register for labour market statistics. Also studies and labour market measures should be included in the register. The labour market is the scene of an encounter between the supply of labour and the demand for it. The supply is a question of persons identified by their civic registration numbers, the demand is a question of local units identified by local unit numbers.

When supply and demand meet, relational objects are created (employment, self-employment, etc.) which are identified by the combination of the civic registration number and local unit number. The Activity Register should be used for new demographic statistics concerning employment etc. Among the users there is a great interest in statistics combining personal data and company data. With the Job Register many such combinations of data are possible, with the Activity Register even more.

Our proposals concerning improvements in Statistics Sweden's present registers are indicated in Chart 5 below. Greater detail is to be found in Section 5.

Chart 5. Registers and linkages which must be improved

Normal = Directly based on administrative data
Italics = Integrater register, no new data

Swedish for Immigrants
Adults Education
Upper Secondary School
Form 9
Teachers
Higher education
Persons Enrolled in Education

Population and Housing Census
Employment

Population register

Activity register

County Councils - wages
Ecclesiastical Districts - wages
Municipalities - wages
Civil Servants - wages

Geographical Database

One or Two Dwelling Buildings
Multi Dwelling Buildings
Industrial Real Estate
Agricultural real Estate
Dwellings

Real Estate register

Business register

Agricultural Register

Register of Schools

The following measures are designed to improve co-ordination within the system:

4a. The basic registers shall be developed into basic registers pure and simple, forming the basis of the system with regard to object definitions and object sets. They shall contain versions differing in respect of time – current version, stock at the turn of the year, change during the year and objects which have existed at some time during the year.

4b. The keys between the basic registers shall be improved. The reference variables in the basic registers shall be standardised and there shall be a high percentage of hits when basic registers are matched.

4c. It shall be possible to link all other registers to their basic register by means of the standardised reference variables. The object sets shall be in tune with their basic register.

4d. The variables in the system's different registers shall be co-ordinated. There shall be no change in the name or definition of the same variable from register to register. Metadata shall be used jointly in order to avoid duplication of work.
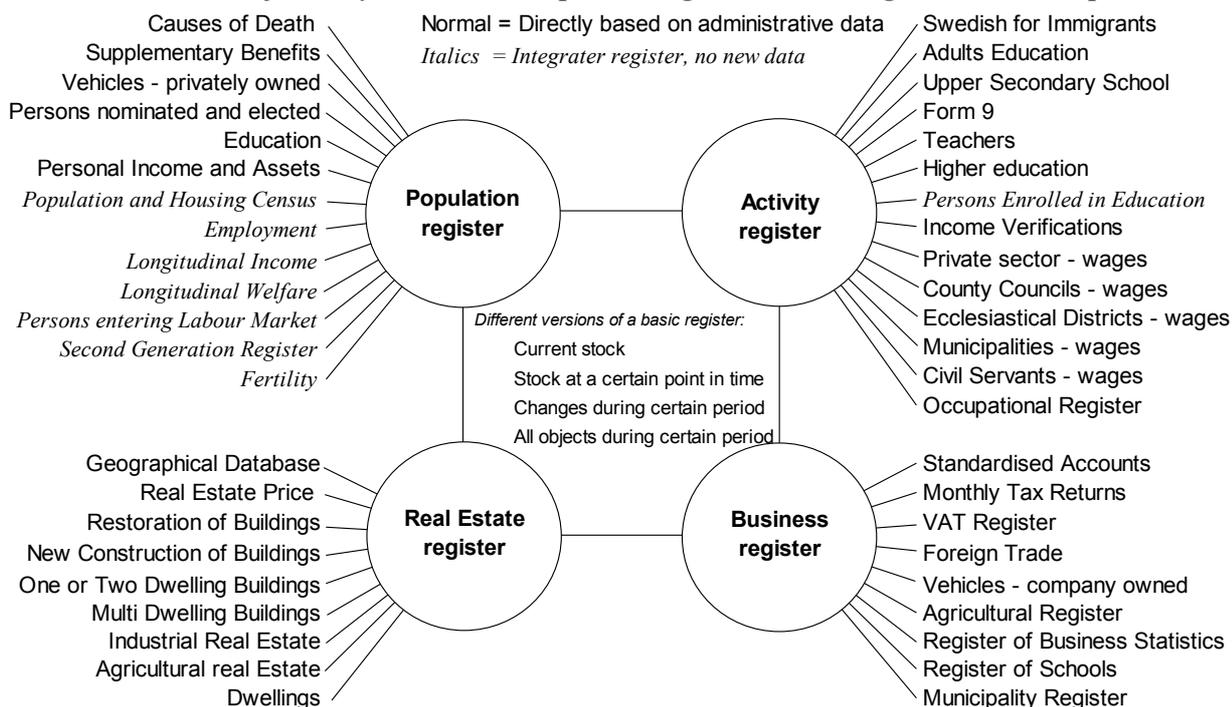
4e. The same statistical methods and IT tools shall be used.

4f. Secrecy and protection of privacy within the system shall be improved.

There is a more detailed account of these proposals in Sections 4–9.

The improvement of existent registers and keys in accordance with our proposals will give our register system the following appearance:

Chart 6. SCB:s future system when our present registers and linkages have been improved

Normal = Directly based on administrative data
Italics = Integrater register, no new data

Causes of Death
Supplementary Benefits
Vehicles - privately owned
Persons nominated and elected
Education
Personal Income and Assets
Population and Housing Census
Employment
Longitudinal Income
Longitudinal Welfare
Persons entering Labour Market
Second Generation Register
Fertility

Population register

Activity register

Swedish for Immigrants
Adults Education
Upper Secondary School
Form 9
Teachers
Higher education
Persons Enrolled in Education
Income Verifications
Private sector - wages
County Councils - wages
Ecclesiastical Districts - wages
Municipalities - wages
Civil Servants - wages
Occupational Register

Different versions of a basic register:
Current stock
Stock at a certain point in time
Changes during certain period
All objects during certain period

Geographical Database
Real Estate Price
Restoration of Buildings
New Construction of Buildings
One or Two Dwelling Buildings
Multi Dwelling Buildings
Industrial Real Estate
Agricultural real Estate
Dwellings

Real Estate register

Business register

Standardised Accounts
Monthly Tax Returns
VAT Register
Foreign Trade
Vehicles - company owned
Agricultural Register
Register of Business Statistics
Register of Schools
Municipality Register

These improvements will make the register system function better and give it greater potential. Existent statistics will be more consistent and of higher quality, at the same time as there will be greater possibilities regarding new register-based statistics.

Old versions of registers shall be filed in the National Archives. For certain commissions Statistics Sweden needs to make use of such registers, and smooth routines need to be established for this.

New integration registers bring certain organisational questions to the fore. To avoid conflict there needs to be regular co-operation between the programs concerned. There shall be no monopoly on the use of certain data. Charges, both internal and external, shall follow the standard and known rules.

5a. In order that all within Statistics Sweden shall be able to exploit the possibilities of the improved system there should be appointed a group with competence from different parts of the register system whose task shall be to develop joint processing and integration registers involving new content. This group will perform advanced development work for the purpose of creating new statistics.

5b. There shall be easy access to registers filed long-term.

5c. There shall be co-operation across program borders when new integration registers are being established.

5d. The rules regarding the pricing of register services shall be clarified and disseminated.


Step 5. Co-ordination of Statistics Sweden's and other authorities' register systems

Up to now we have only discussed Statistics Sweden's register system. The registers that make up this system can be both ones that Statistics Sweden has the responsibility for and ones that other authorities are responsible for. Furthermore there are many external registers which could be linked to Statistics Sweden's basic registers.

Step 4 above, involving the creation of Statistics Sweden's register system as a totality, presupposes increased co-operation between departments and programs within Statistics Sweden. Similarly, Statistics Sweden and other authorities with responsibility for statistics should co-operate concerning register questions in order to establish a register system for the entire statistical system.

It is important that Statistics Sweden have access to good registers from other authorities, and that these authorities in turn have access to Statistics Sweden's register system. Thus co-operation with regard to register questions is of mutual interest.

Chart 7. The register system of Statistics Sweden's and other statistical authorities
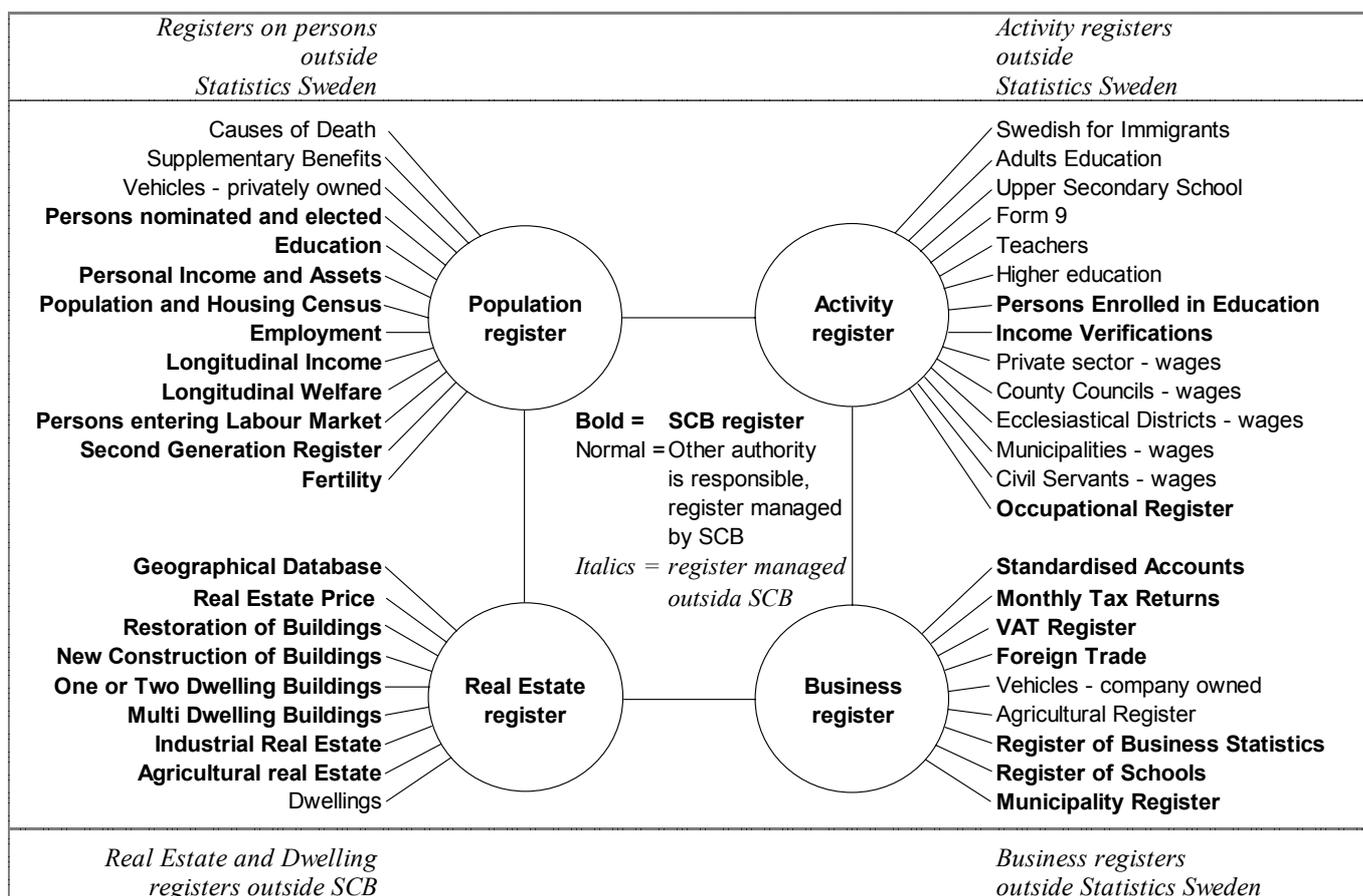
The Vehicle Register (covering both private and organisation-owned vehicles) is used for Statistics Sweden's national accounting and could be used in other contexts, e.g. that of welfare statistics. By linking car-owner to education, income, etc. from Statistics Sweden's system the Institute for Transport and Communication Analysis could produce new statistics of its own.

The different education and salary registers will play a more important part in Statistics Sweden's statistics if our proposals are followed. We believe that they should be used for fulfilling the requirements of the Activity and Employment Registers, whereby they would have a beneficial effect on the quality of future register-based population and housing censuses.

To enable the Farm Register to be linked to the Business Register the unit concepts must be in accord. Thereby the method of defining agricultural enterprises is of importance not only for the Board of Agriculture but also for Statistics Sweden and the entire statistical system.

> *6. Statistics Sweden shall devise a strategy for co-operation concerning register questions with all the authorities concerned.*

4. Co-ordination and standardisation of variables

At present Statistics Sweden's store of data is strongly decentralised, both physically and organisationally, with more than 1,000 databases stored on some 60 servers. A prerequisite for the efficient production of statistics is that the registers in the store are co-ordinated both technically and with regard to content. This makes it possible to avoid unnecessary special solutions – solutions that can cause confusion when data are interpreted, matched or compared. When it comes to content it is important that there be co-ordination both of variables and of classifications. An important tool of co-ordination is the Classification Database, which contains statistical classifications and is used within Statistics Sweden and the Swedish statistical system. It contains classifications and keys which have been established as standard in Sweden or internationally. An example of such a classification is branch of the economy, indicated in accordance with the Swedish Standard of Industrial Classification of All Economic Activities.

No great effort has hitherto been made to create uniformity among the variables in different Statistics Sweden registers. The change of platform involved extensive reorganisation and a review of production, and there is little doubt that Statistics Sweden here missed a fine opportunity to enact a co-ordination of variables. The lack of documentation has earlier meant that it has been difficult to compare different definitions and descriptions of variables. There are examples of variables with the same content but with different names in different registers, as also of variables with different definitions but the same name.

The situation today

In 1999 we began an initial inventory of variables on the basis of the Population Register. Comparisons were made with variables in a number of other registers, and differences emerged. A proposal regarding the co-ordination of variables was drawn up, and on 3 April 2000 came a decision by the Director-General of Statistics Sweden to the effect that certain important variables should be standardised.

These variables are such as are frequently used in the statistics production other registers. The register that takes in a variable from another authority or creates a variable of its own shall be responsible for the said variable at Statistics Sweden.

We have made a survey of variables in Statistics Sweden's basic registers which are suitable for standardisation. This was done in co-operation with representatives of the registers concerned. The work of co-ordination began with a review of the variables in the Population Register, six of which were then initially proposed as appropriate for standardisation: civic registration number, gender, age, civil status, country of birth and citizenship. We would say that there are about ten variables in the basic registers that should be standardised to start with.

We have also reviewed the list of 13 central reference variables, since these also are a type of standardised variables. In order that the list shall be in line with the above-mentioned decision the variables included should be such as are used by several registers. From contact with the department programs concerned it has emerged that at the moment there are also a number of variables on the list which are only used within the particular program itself or which are no longer of interest. When it comes to these variables it is not the head of department that makes the decision, since the original list has emerged in the form of a decision by the Director-General. Even if the reference variables are not

changed, information about them should be distributed to the staff, because we have found that not everyone knows of their existence.

It needs to be investigated what technical solution is suitable for making the standardised variables accessible. Earlier the Classification Database included a section including reference variables. This database is at present undergoing new development, and no variable section has been planned for. Nevertheless a possible long-term solution is to put variables into the Classification Database, this because of the close connection between variables and classifications. Alternatively there could be a separate database with linkage to the classifications in the Classification Database. Another feasible solution is a search function in Metadok. As a first step the variables in question, together with the pertinent information, could be brought together on a simple list, which ought to be adequate until a greater number of variables have been selected.

When it comes to the creation of completely new registers it is easy to follow norms, but in order to achieve full co-ordination it must be possible to change variable names and codes, for instance, in existent registers and in the procedures used for the creation of new sets of annual statistics. For this reason it is difficult to set forth a realistic timetable for carrying out these changes.

> *7. The standardisation of variables calls for investigations of a technical kind. A good technical solution is needed such as to make metadata for the standardised variables easily accessible. Furthermore there needs to be an appraisal of how changes in variable names, codes etc. can be introduced into production.*

Measures proposed

For a co-ordination of variables in accordance with the above-mentioned decision of the Director-General to be carried out, a central function is needed. As things appear at present there is no natural place in the organisation for this type of work, nor has any individual program sufficient resources to handle it alone.

> *8. A central register function is described in Section 10, and the area of responsibility of such a function should include advancement of the work of standardisation and the formulation of clear rules and instructions regarding how the standardised variables are to be managed and followed.*

It shall be the task of staff working on the register responsible to make an inventory of the variables to be standardised and to appoint a person with responsibility for them. This inventory shall then be approved by the central function. A definite decision regarding the variables in question will thereafter be made by the head of department concerned. The responsibility of the aforesaid person concerning the variables covers both documentation and information. What the documentation responsibility implies is that the variables shall be documented in Metadok in accordance with instructions and copied to the central database. What the information responsibility implies is passing information to, and consulting with, staff on registers that regularly fetch a variable if a change is planned. Information shall also be passed to the registers concerned if an error is discovered in data/metadata.

> *9. In order to achieve increased co-ordination with regard to content, variables in the basic registers shall be standardised and the central reference variables be updated. As a second step there should be a review of variables in other registers. We suggest that the central register function be given the overall responsibility for the carrying out of this.*

5.   Methodological work at the system level

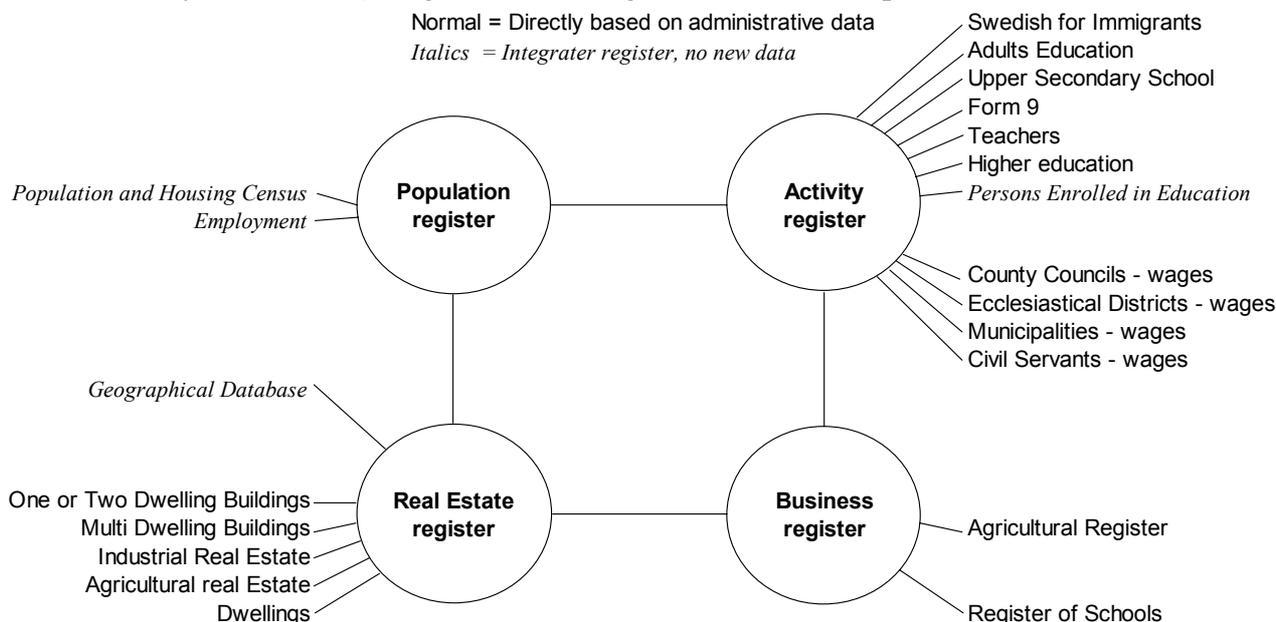Register-statistical methodological work can be pursued on two levels:

– Methodological work at the register level. This work involves responsibility for the methods employed when a statistical register is created. (There is a discussion of it in Section 6.)

– Methodological work at the system level. This work involves co-ordinating a large number of registers into a functioning system of statistical registers. The system as a whole shall be developed in order to make possible new and improved statistics.

Even though the methodological work at the register level is pursued within the particular register programs there must be co-ordination so that all registers can be used together. The methodological work at the system level presupposes knowledge of the whole system and thus cannot be replaced by such work at the program level.

---
10. The overall responsibility for the methodological work at the system level should be at the centre – it should rest with the central register function, we would suggest. There is a discussion of this in Section 10. There must be continuous contacts between this central function and the different register programs. The methodology groups of the departments must contain at least one methods statistician with responsibility for register questions.

Together with the contact group we have studied the different parts of the system. We have also made an inventory of the different registers included in the system. On the basis of our knowledge of the system as a whole we can suggest a number of ways in which registers can be improved. This was mentioned in Section 3, where the proposals were summarised in the following chart:

Chart 5 (from Section 3). Registers and linkages which must be improved

Normal = Directly based on administrative data
Italics = Integrater register, no new data

Swedish for Immigrants
Adults Education
Upper Secondary School
Form 9
Teachers
Higher education
Persons Enrolled in Education

County Councils - wages
Ecclesiastical Districts - wages
Municipalities - wages
Civil Servants - wages

Population and Housing Census
Employment

Population register

Activity register

Geographical Database

One or Two Dwelling Buildings
Multi Dwelling Buildings
Industrial Real Estate
Agricultural real Estate
Dwellings

Real Estate register

Business register

Agricultural Register

Register of Schools

The system of basic registers

The system of basic registers forms the basis of the entire register system and consists of four registers containing important types of object. If any of these basic registers should be missing or be of low quality the entire register system would be considerably less utilisable for statistical purposes.

For Statistics Sweden's registers to function as a system it must be possible to link the objects in different registers by means of reference variables. An efficient system requires that these variables be of good quality and that the same

ones be found in several registers. The keys between the basic registers are of particular importance.

The Population Register

This basic register has advanced furthest when it comes to the development of register-statistical methodology for work with a basic register. The persons working on it know how to update, have a long tradition of producing advanced register statistics, regularly perform demographic analyses of the object population and are accustomed to creating integration registers for different purposes. They have an advanced production system where the history of the objects is stored. The Population Register can serve as a source of inspiration for the other basic registers. Nevertheless there are certain negative aspects:

– The household definition have faults at present. A good register of dwellings can remedy this.

– There is overlap in the coverage of certain categories of immigrant. This was discovered within Statistics Sweden, and the National Tax Board was informed. The board is now improving its check on population registration regarding these immigrant categories in order to get rid of the overlap.

11a. It can happen that a civic number is used again (the same number being used for more than one person, that is) and that a person's civic registration number is changed. This presents certain matching problems which could be avoided if Statistics Sweden serial numbers were introduced in the same way as has been done in Denmark. In Section 9 we suggest that serial numbers be introduced in order to improve the protection of privacy, and the suggestion is given added force by its constituting an answer to these matching problems.

11b. There should be an annual register with all the persons who have been registered during a specific year. When different registers are matched they must all have reference to the same point in time/period. Both the Activity Register and the Income Verification Register have reference to calendar years and need to be matched with such an annual version of the Population Register.

11c. Reference variables should be information-free so that for instance a change in regional division does not affect the reference variable.

11d. At present county, municipality, parish and real estate designation are used as reference variables vis-à-vis the Real Estate Register. Instead of matching with these four variables there should be a particular national real estate number.

The Real Estate Register

The present Real Estate Register does not have the same structure as the other three basic registers. With its hundreds of statistical variables this register is too rich in content to be a basic register pure and simple, and it is updated only once a year. The linkage to the Business Register is poor.

12a. There should be a straightforward basic register with reference to real estate and related objects (owners, assessed units, valuation units, buildings/land). The co-operation between the National Tax Board and the National Land Survey Administration should be developed so that Statistics Sweden will get continuous updates to this new basic register.

12b. The key between the Business Register and the Real Estate Register needs to be improved. The geography database must be developed so that the addresses of the local units can be connected to the right properties. At the Land Survey Administration there is an awareness of the importance of business addresses, since these are needed for studies of local and regional development.

The Business Register

Statistics Sweden receives data concerning legal units from the Patent and Registration Office and the National Tax Board. Statistics Sweden devotes wide-ranging effort to collecting information from companies with operations at more than one work-place, in order to create local units. This work is divided up between two departments. The department of labour market statistics works on income verification statements and is frequently in contact with companies when such statements are being edited. Such statements offer an example of integrated data collection, where not only tax-administrative data are collected but also certain data for statistical purposes (e.g. the demarcation of work-places). This work has to do with the work-place population for the previous calendar year, in contrast with the register maintenance work within the department of economic statistics, which is concerned with the current stock of work-places. Here, too, multiple work-place inquiries are carried out and there is frequent contacting of companies. Within the Business Register are also created company, activity and local activity units.

Two principles permeated the work on development of the Business Register:

–   No parallel object sets shall appear within economic statistics. Thus for instance the population of industrial enterprises shall be the same in industrial statistics and the Business Register.

–   Everyone within the department of economic statistics shall participate in maintenance of the Business Register. Those who for instance collect data from industrial enterprises shall pass on the information they receive about the industrial population to the Business Register.

We consider that these principles should be adhered to throughout Statistics Sweden, not just within the department of economic statistics. This means that, for instance, those working on the Farm Register shall provide the basis for the agricultural part of the Business Register, and that those working on registers concerning schools and other places of education shall provide the requisite data for making the educational part of the Business Register as good as possible. The same applies to the other three basic registers – there must be no parallel object sets!

13a. Annual versions and stock registers with reference to the turn of the year should be regularly created. Both the Activity Register and the Income Verification Statement Register refer to the calendar year, and it needs to be possible to match them with such an annual version of the Business Register. Stocks at different turns of the year and changes during the particular years are required for demographic business statistics.

13b. There should be the same object definitions and reference variables regarding companies, business units and local units in the Business Register as in the Activity Register. The Business Register reflects current stock. The Activity Register, based as it is on income verification, provides partially conflicting information about the same population after long delay.

13c. Concerning certain parts of the business population the present-day Business Register has unnecessary deficiencies of quality. Current information concerning agricultural enterprises and places of education is to be found within Statistics Sweden but is not made use of by the Business Register. If corporate registration numbers and local unit numbers are introduced in agricultural and educational statistics, respectively, the quality of the Business Register will benefit.

13d. No parallel object sets shall appear. This means that all registers must be in agreement with the relevant basic register.

The Activity Register

The current Job Register constitutes the foundation of the register-based employment statistics, which are a basic prerequisite for the carrying out of a register-based population and housing census. The register at present describes different forms of gainful employment, including self-employment. The register would have a better content if the alternatives to gainful employment were added. Work has begun on the creation of such a basic register for labour market statistics. This Activity Register will cover not only gainful employment but also such things as studying, looking for work, being off sick and being on disability pension. It is important that the work on this register be completed. Reference variables in the Activity Register are civic registration number, local unit number and corporate registration number.

Whilst the traditional form of presentation offers too static a picture of change in society, the Activity Register gives us the chance to obtain a more dynamic picture. Preliminary studies of the Register Project's test database (which is a miniature version of Statistics Sweden's register system) indicate that about a third of gainful employment activities disappear within a year and are replaced with new ones.

If also study activities are to be linked to the Activity Register the place of education must be identified by a local unit number and the arranger of the education by a corporate registration number. This will also give better commuting statistics, describing the commuting both of persons gainfully employed and of persons studying.

The employment register, which at present is based on the Job Register, suffers from certain errors of classification. An evaluation indicated that 5% of employed persons were wrongly classified as non-employed and that 14% of non-employed persons were wrongly classified as employed. It is November employment which is to be described, and as the time specifications on the income verification statements are not of a high standard we find ourselves with these errors of classification.

The salary registers, on the other hand, contain information about November employment (in certain cases September–October) with precise and reliable indications of time. Linking these registers to the Activity Register would greatly improve the quality of the employment register. This will be important for the 2005 population and housing census.

14a. The Job Register shall be expanded into an Activity Register which also includes studying, looking for work, being off work a long time because of illness, early retirement and disability pension, etc.

14b. The school and higher education registers shall have keys to the Activity Register by way of the reference variables civic registration number and local unit number.

14c. The salary registers shall have keys to the Activity Register in the same way.

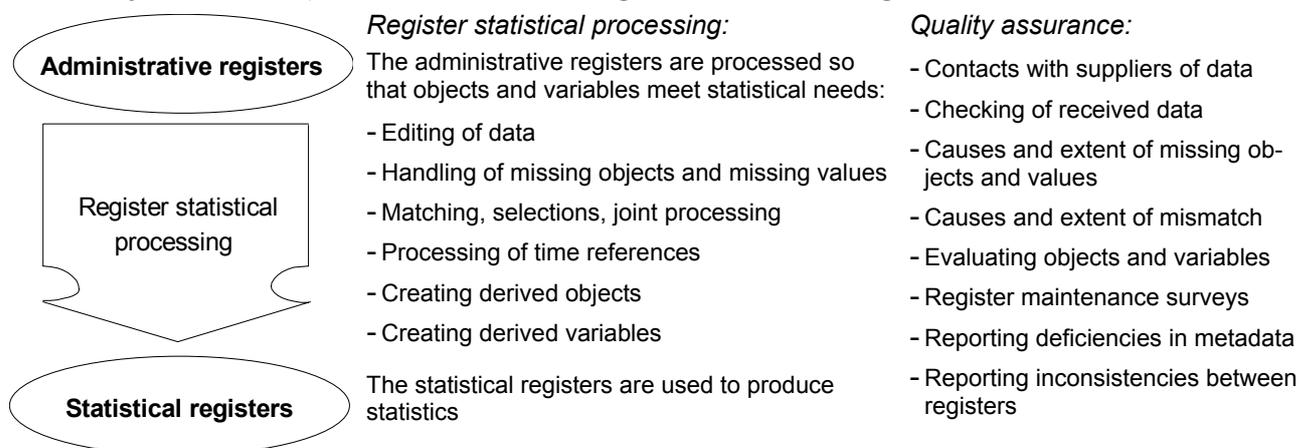# 6. Methodological work and quality assurance at the register level

Statistics Sweden's sample surveys employ well-documented methods which are discussed in reports and at seminars. Many highly qualified statisticians work on these surveys. When it comes to register statistics, however, the situation is different. Ad hoc methods are employed, these methods are not discussed and have not been documented, and few statisticians have taken an interest in this field.

One might say that half of Statistics Sweden's operations have been – and still are – invisible. When questions of method are discussed, the sample surveys command the stage. Such dominance is quite natural when it comes to questions of sampling and estimation, less so when it comes to editing, non-response and documentation in SCBDOK. In Statistics Sweden's publication Att köpa en statistisk undersökning (Buying a Statistical Survey) only sample surveys are mentioned, this in spite of the fact that many of the surveys commissioned are register-based. To some extent the work of the Register Project has changed this view of things, but much remains to be done.

> *15. The register-statistical work shall be given the attention it deserves. Register-statistical methods shall be discussed and documented. Register processing shall be based on established methods of proper quality.*

In Section 3 the different components of register-statistical work were presented.

Chart 2 (from Section 3). From administrative registers to statistical registers

	Register statistical processing:	Quality assurance:
Administrative registers	The administrative registers are processed so that objects and variables meet statistical needs:	- Contacts with suppliers of data
		- Checking of received data
	- Editing of data	- Causes and extent of missing objects and values
Register statistical processing	- Handling of missing objects and missing values	
	- Matching, selections, joint processing	- Causes and extent of mismatch
	- Processing of time references	- Evaluating objects and variables
	- Creating derived objects	- Register maintenance surveys
	- Creating derived variables	- Reporting deficiencies in metadata
Statistical registers	The statistical registers are used to produce statistics	- Reporting inconsistencies between registers

Administrative data must be processed so as to meet statistical requirements. It is far from evident that there is adequate awareness of this. In Register Project contacts with non-Nordic countries the mention of such processing has caused something of a stir. There is scepticism regarding the use of administrative data for statistical purposes. We believe that this scepticism is attributable to certain misconceptions:

– That use is made of administrative data as they are, without processing and quality assurance. Chart 2 above shows that this is not the case.

– Only one administrative source is used for a certain purpose. If the source is part of a system of registers several sources will be compared and each of them will contribute to raising the quality of the system.

Incorrect and uncritical use of administrative data does occur, but it can be prevented by means of metadata providing information about problems of comparability. Metadata have an extremely important role when it comes to working on register statistics. When one creates integration registers where data from

different registers and years are processed together, there is great risk of serious error if one is unaware of problems of comparability and changes of definition.

Methods of processing

In the same way as we have made an inventory of the registers within Statistics Sweden, we need to make an inventory of the methods used for the creation of statistical registers. A natural task for a central register function is to proceed with this inventory and provide methodological support for the register programs.

Check of basic data

Statistics Sweden has carried out a major editing project. The report from the editing group (March 1997) gives an account of methods for editing in the first place sample surveys. Macro-editing, which is the method recommended in the report, is probably less suitable for register statistics.

There must be an account of the editing, describing on the one hand the method used, on the other hand the extent of different types of error and the measures taken to rectify them. The register-supplying authority should receive feedback when errors are discovered

Supplier contacts

Co-operation between Statistics Sweden and the authorities which supply administrative registers must be a regular feature of the work on register statistics. Statistics Sweden must have full information about the rules and concepts used in the administrative systems – must have access, that is, to administrative metadata. Statistics Sweden should furthermore be given the opportunity to discuss deficiencies and to offer views on changes planned in the administrative systems, so as to obtain in good time information about system changes which may affect the content and quality of the statistics. Certain changes may call for extra measures on the part of Statistics Sweden in order to ensure the quality of the statistics – and Statistics Sweden must get the authorities to see this.

Measures occasioned by missing objects and observations

Here, too, Statistics Sweden's thinking is dominated by non-response problems in the case of sample surveys. In the annual Non-response Barometer it is only the non-response within sample surveys that is taken account of. Also the non-response manuals which have been produced are wholly directed towards the problems arising in sample surveys. For the variables in the system the extent of partial non-response should be documented, as also should be the methods of non-response compensation.

Joint processing and selecting

The joint processing of registers calls for knowledge and experience to avoid misleading results caused by insufficient comparability or other types of quality problem. Mismatch between the basic registers must be documented, as must mismatch between a particular statistical register and its basic register. The causes of mismatch must be gone into and the results documented. Mismatch might be caused, for instance, by object drop-out or defective reference variables in the case of one or more of the registers involved.

Reporting of inconsistencies between registers

All who are involved in the joint processing of registers make valuable observations which must be documented and reported to the persons responsible for the registers in question. It needs to be easy to report any deficiencies or oddities discovered, and those responsible for the registers must investigate the causes. The results of this are then to be put into the metadata system.

Evaluation of the quality of objects and variables

The sample surveys which are regularly carried out should be used for evaluation of the quality of objects and variables in the register system. It is particularly important that derived objects and variables be regularly checked. Special questionnaires for register maintenance must be used in certain situations.

Quality assurance [1]

The traditional error model can be seen as following the different steps of a sample survey. The errors which appear during the process are these: relevance error, framework error, non-response error, measurement error, processing error and sampling error.

In a similar way we should like to describe the quality of register statistics by following the different stages of production given in Chart 2 above. Quality assurance involves investigating and remedying deficiencies of quality which may arise in different parts of the work with register statistics. All such deficiencies must be documented and the relevant information passed back to those concerned.

There is at least as much need for quality work, methodological studies and quality assurance in the case of register-based surveys as in the case of sample surveys. Where the latter are concerned the producer can by means of his/her own measures and choice of methods set the level of quality and measure the quality of the published statistics quantitatively by estimating margins of error for the stochastic sampling errors.

In the case of register-based surveys the situation is completely different: the person responsible for the register cannot control the data collection and does not know how the register is going to be used or what estimates are going to be made by persons within and outside Statistics Sweden. Furthermore it is difficult to devise quantitative measures of quality, and there is no parallel to the well-established methodology for calculating margins of error in sample surveys.

Often, however, there is a qualitative knowledge in that the person responsible for the register is well aware – through his/her observations, experience and knowledge of the subject – of the occurrence of a number of types of error even though he/she is unable to measure how common or significant they are. This less precise type of knowledge should also be documented – it can warn users against uncritical use and can indicate a need for improvement. Knowledge of reliability can be classified in accordance with its acuity on one of the following levels:

1. No knowledge at all.

2. Vague qualitative information based on judgment.

3. Documentation of the production processes and their characteristics.

4. Systematic qualitative information based on surveys.

5. Quantitative indicators of reliability, e.g. extent of mismatch.

6. Quantitative measures of the reliability of the estimate.

---

[1] This section is based on Håkan Lindström: Kvalitetssäkring i register för statistikproduktion med administrativt underlag (Quality assurance in registers for statistics production on an administrative basis). Report from the Register Project, November 1999.

2 (From Section 3). The receiving of large volumes of administrative material and the editing of this material are important operations that many programs perform in the same way. These operations should be made more efficient by means of exchange of experience and the development of common tools and methods of editing. Similarly, methods and tools should be developed for non-response management, joint processing and evaluation. The responsibility for this should rest within the general methods program and the work should be carried out in close con-tact with the register programs.

16a. Methodology and tools for creating integration registers shall be developed on a common basis. The deficiencies of quality discovered in the course of joint processing shall be reported to the registers concerned.

16b. The basic registers share problems of statistical method and IT problems. For this reason they should co-operate with regard to such questions as the evaluation of object quality, coverage, updating, demographic methods, system structure and IT tools.

7. Production process and operational environment

In order that the register system shall be able to function as a strategic resource at Statistics Sweden it is important that both the design of the production process and the operational environment should support such a development. The new IT platform does not automatically give this support. In the final report of the Platform Project it is stated that there has been considerable professional development but the technical platform needs further development and completion to make the production process more efficient. There is a lack of uniformity when it comes to system solutions, software environment and the array of IT tools. It is also stated that register-based production and a well-functioning metadata system are going to continue to be of great importance.

The observations and experiences concerning the production process which are described in this section support the above. Within the project we have on the one hand acquired our own practical experience of register processing, on the other hand discussed these questions with persons involved in the Platform Project and the Metadata Project. The contact group have often emphasised problems to do with production. Statistics Sweden must concentrate on solving these problems if it is to be possible to fulfil the goal of having an efficient and co-ordinated register system.

The further development of the register system requires a well-functioning operational environment and production process. It must be possible to process several registers jointly even if they are large. Furthermore metadata should be made available at the production stage.

The Bamse example

The problems that exist today can be illustrated by the state of affairs concerning the so-called Bamse server. Since the late 70s Statistics Sweden has been developing a special service for municipalities in the form of basic data at the sub-area level. The system has been gradually extended from comprising principally population and census material to including also data from the vehicle register, the employment register, the income register, etc. In the new environment it was necessary to collect this heavy processing and the associated commissioned services in one server. In line with the carrying out of the platform change the latter server has become heavily burdened, and often 80–100 sessions are going on at the same time. The processing can take a long time, therefore it is difficult to draw up a schedule and promise customers a time for delivery. It is worth noting that the IT Unit and the affected IT groups at the departments have for some time been making a great effort to solve the problems and thereby lay the foundations for a functioning operational environment.

The fact that the problems regarding the new platform have been taken up chiefly by the staff concerned with Bamse is probably attributable to Bamse's involving joint processing based on customer requirements and to this involving the processing of many large registers. In the light of this it is fairly natural that the most advanced demands concerning the processes and the operational environment should come from persons working with the Bamse server. Probably in the future similar problems will appear in other parts of the register system.

Metadata and documentation are also needed in the process

Metadata and documentation do not constitute a particular part of the production process coming at the end but are created, ideally, "en route". Unfortunately, however, we have found that it appears to be difficult to achieve the desirable co-ordination, not merely between the different parts of the metadata system but also between the metadata system and the different production sys-

tems, despite the fact that there is a good theoretical base and that the different parts rest largely on the same conceptual and descriptive model.

For many users to be able to make efficient use of the register system, metadata must be built in. Unfortunately there are few tools that can handle metadata parallel with data. The idea is that metadata for the registers to be jointly processed shall be included in the case of matching, selection and aggregation. A matching results in a new, temporary register that contains precisely those variables required for the current task. Initially there are no metadata connected to this register. Here it should be possible to easily download metadata from the registers that are matched and link them to the temporary register by means of, for instance, a PC Axis. This calls for the possibility of importing existent metadata from Metadok.

The need for metadata is also evident when it comes to generating a large number of tables with numerous divisions and many variable values that need to be written in. Here again, a tool which by means of which this need could be fulfilled, subject to certain development, is the PC Axis. Tests have shown that the program can cope with pure SQL questions, the handling of which requires far more training. The PC Axis has been chosen as the standard format for the delivery of standardised regional tables to municipalities by way of Statistics Sweden's website, and its structure is suited to the management of metadata. Functions already exist for the grouping of hierarchical variables, and tables can easily be exported to other formats.

Production methods and tools

It has been a natural goal for Statistics Sweden to build up a general data store available to "all". Within the framework of the Platform Project and the Register Project there emerged a vision of this store, thereafter presented at a meeting of managers. Yet there are still no clear strategies or methods for the realisation of the vision. As pointed out in an analysis of the Bamse situation, databases vary greatly in design, many of them having been designed to meet the requirements of a particular program and without much concern as to how databases are to be co-ordinated. Furthermore, the structuring of input and output databases is commonly indicated as being an important efficiency and quality question – but how this structuring shall be carried out does not appear to have been adequately gone into.

> *17. It is important that adequate resources be allocated to the measures now planned within IT regarding a development of Statistics Sweden's data store and the management of large volumes of data.*

There is need of a general matching tool for the new SQL environment. Persons in production need to be able to easily match registers so that there shall emerge on the one hand an SQL table with the appropriate variables for the objects in all the registers involved, on the other hand different SQL tables with certain variables for different categories of non-matching objects. There are examples of menu-based modules where the matching is described graphically and then generates the necessary SQL code.

Not only would a general matching tool facilitate production, it would also mean better quality in that on the one hand there would be less risk of error in matching, on the other hand the tables created in the case of mismatch could be used for further analysis of the non-matching objects.

In 1999 there was a full-scale experiment with a multi-server join and large volumes of data, and from this came the following conclusions: (1) it is possible to perform large-scale multi-server processing, (2) servers should be structured in accordance with type of processing, and (3) is worth using a multi-server join in some cases instead of double-storing data. It must be emphasised,

however, that with a multi-server join the processing times were much longer than they would have been if there had been a corresponding join within one and the same server. This is probably why no one – or at least hardly anyone – in fact makes use of the technique.

Statistics Sweden's production environment is all too diversified when it comes to the tools used, which is why there are an unnecessarily large number of system solutions (SQL, SuperCross, PC Axis, Excel, SAS, SPSS). The PC Axis has proved useful for the aggregation etc. of microdata as well as macrodata, but a recent evaluation indicated that this is more or less unknown within Statistics Sweden. AGDA, developed by Statistics Sweden for aggregation, has only four users, whilst GRETA, developed for editing, has only one user.

18a. The question of the production processes of the future and the requirements regarding the operational environment should be studied and resolved with particular attention to the problems which have emerged in the case of Bamse. There is a risk that the Bamse problems may be regarded as an exception – within an environment which otherwise functions rather well – rather than as ones which need to be solved if and when the entire register system is to function full-scale.

18b. The production environment is staff-dependent and not greatly standardised. General tools are lacking (or are too many in number to be maintained) and it is unfortunate that so much programming competence is required for performing the most common tasks. The new IT organisation needs to acquire a better grasp of the whole, to analyse the process and to thin out the flora of tools, taking clear-cut measures.

Exchange of experience and professional development

In the form of the new IT organisation there has been created a foundation for greater co-operation between different parts of the broader organisation. Thereby more persons have access to a common basis for system design and to practical professional development. It is a matter of urgency that this should go through also with regard to the necessary co-operation when it comes to work on register-based production. We have found great variation in the practical methods employed in the register process. Good solutions must be disseminated and put to use, and shared assessments can form a basis for the development of general tools as a support for the work. It is important that this should not just involve IT staff but be extended so as to include persons who are responsible for registers and who with more standardised aids would themselves be able to assume responsibility for a greater part of the production than now.

---

19. From the Platform Project there emerged a proposal for "experience seminars" based on different production stages and products. This proposal should be acted upon by the new IT organisation.

8. Documentation and metadata

Simplifying, one can say that metadata are data about data. Metadata describe the content and technical structure of a body of data and are important for interpreting the signification, possibilities and limitations of microdata. Therefore metadata are of great importance for both internal and external users. Register statistics are hampered if there is little access to micro-metadata, therefore we within the project have made a great effort to foster more active documentation. One expression of this effort was that a number of important registers were collected in a particular project, the Documentation Project, to help the staff working with these registers to set about the task of documentation.

In this connection it was natural to think first of the internal users who need metadata when they are working on the production of register statistics from their own registers and other registers within Statistics Sweden. Metadata are a support for the producer of statistics when it comes to maintenance and development of the production system. Good access to metadata makes production more efficient and less vulnerable to staff changes, at the same time as it facilitates the training of new staff.

Metadata are also of great importance for our external users, enabling them to search among the available registers and make correct interpretations and analyses of the statistics they receive from us. Furthermore metadata facilitate communication between customer and producer with regard to the ordering and production of tasks. Thus also the needs of external users call for a well-functioning metadata system.

The situation today

At Statistics Sweden there are a large number of registers which are completely enumerated. A prerequisite for their functioning in a register system, the way we see it, is that there are sufficient metadata regarding each one of them. This calls for extensive documentation work at the departments. The responsibility for documentation of the registers rests with the respective programs. This has in fact been in the department guidelines for several years, but things have gone slowly.

On the basis of an inventory of the documentation situation regarding the registers in the register system which was made at the turn of the year 2000/2001 it can be stated, we are happy to say, that almost 70% of the registers have at least one annual version documented in Metadok, whilst many registers have several such versions completed. The situation is less rosy when it comes to SCBDOK. Here the proportion of registers with at least one such documentation completed is only about 30%, and in the majority of cases only one annual version is documented. Unfortunately many of these documentations have not been sent in for editing and putting on to the Internet.

There are several reasons why the work has not been performed at the desirable pace. Documentation work had to be put on the back burner when the platform change called for a large labour input. At the project's contact group meetings a number of persons have also pointed out that whilst the register staff are faced with the demand for documentation not enough has been done in the metadata area. For lack of resources (both staff and money) the program development of Metadok, for instance, has gone far too slowly. Furthermore many persons have encountered technical difficulties when they have used the program. One reason why the number of completed SCBDOK documentations is low would appear to be that many persons have difficulty in understanding how the different sections are to be filled in. This has often come up at the contact group meetings.

Metadata and the documentation system have during recent years been developed principally in response to the demand for publication of the official statistics in the databases. The work has borne the imprint of this perspective, which has meant that completed surveys and microdata registers have been documented in order to make the statistics more easily accessible for the external users. We should like, though, to emphasise the important internal role of documentation. A new phase must now be introduced where metadata and documentation are created and used successively during the course of production, resulting in the necessary final documentation. This places new demands on the metadata model and the development of tools. Ranges of values which have been put into Metadok, for instance, should also be able to be used in the production of statistical tables instead of there being special auxiliary tables put into the particular database.

An important component of the metadata system is the Classification Database. For quite some time it has not been accessible on Statistics Sweden's network because of reorganisation, and many have found that this has hampered their work. There is a desire that more classifications be put in and that updating be done faster. This would improve co-ordination and facilitate work on the programs, preventing the creation of closely similar groupings into categories of the same variable in different parts of Statistics Sweden.

During the course of our work it has also emerged that there is sometimes inadequate co-ordination between different projects that have to do with metadata. Despite there being links between the projects no one has had overall responsibility for co-ordinating the work.

Proposals regarding measures to be taken

Documentation often has to cede precedence to other tasks which have a higher priority on the programs. To get this work properly going it is important that the following up of the signals from the directorate should be more evident. There could be, for instance, regular reporting of the documentation situation regarding microdata.

For it to be reasonable to place demands on the department programs with regard to documentation, more resources must be allocated to this area. Metadok's functioning must be stable and new functions which many persons call for should be developed faster. In the Metadata Project which has had responsibility for the further development of the program there was until recently only one person working on this. Since autumn 2000 there has also been a back-up and this has meant a reduction in vulnerability, but there need to be more programmer efforts in order to speed up development. The list of desiderata regarding new development is already long. There is for instance a call for a possibility of documenting longitudinally and searching by way of variable instead of by way of register or year. There is also a call for a link to Power Designer and SAS, so that existent metadata can be put to use. We ourselves, on the Register Project, would like to see quality information accorded more space in Metadok and would welcome the development of a general calendar where internal and external events which affect statistics can be inserted.

It is also important, of course, that there should be resources for user support and editing. Completed documentations need to be put on the Internet after editing by the central IP database administration. Lack of staff for such important projects as the Metadata Project, the Database Project and the Register Project has meant that it has not been possible to do the editing at the desirable speed. There is an urgent need to get rid of this bottle-neck so that those who document can see that the material is put to use. If it takes too long a time for the submitted documentation to come into use it is difficult to motivate staff to spend time on such work.

> *20. More resources are needed in order to facilitate the work of documentation for the department programs which are ultimately responsible for the documentation. Resources must also be allocated to user support and the editing of completed documentations. At the same time there must be a central follow-up of the documentation situation.*

The basic idea when it comes to the metadata system is that once information has been put into the system it shall be able to be used in several ways. In cases where the same information is requested in several places, as in "Description of the statistics", quality descriptions, SCBDOK and Metadok, it shall only be necessary to insert the requisite information at one place, this input then being available for use in the rest of the system too. It can be difficult to grasp how the different parts of the system cohere; therefore a general manual should be produced to assist the user in this respect. Metadata, for instance, which have been put into SCBDOK and Metadok can be used for filing, whilst Metadok documentation is needed for handling enquiries concerning the Data Act.

Chart 8. Statistics Sweden's system of metadata
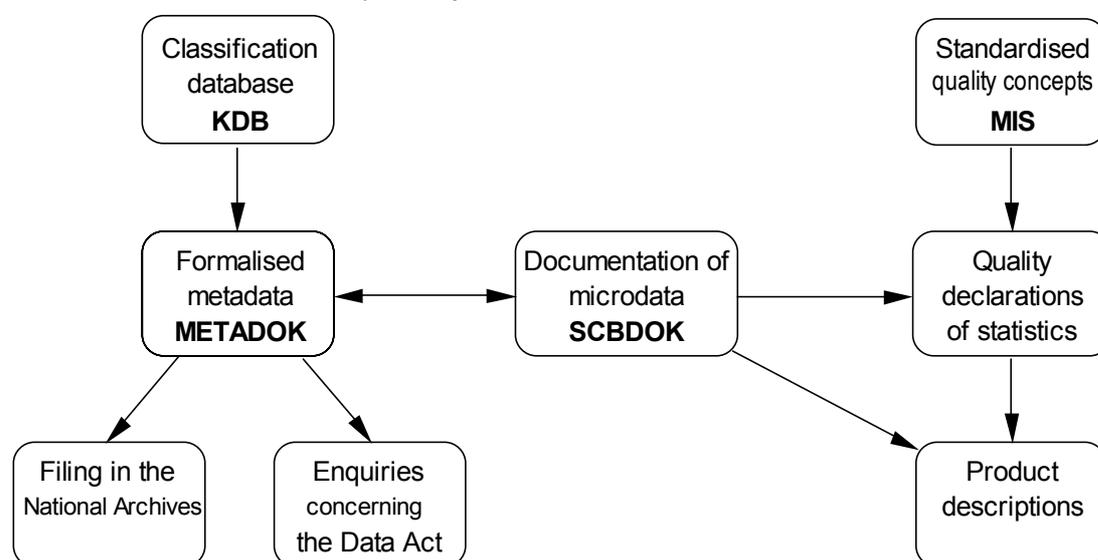
The Register Project's contact group meetings have been an important and appreciated opportunity for discussing metadata questions. At these meetings information has been given regarding plans and current work within the Metadata Project, and the users have been able to put forward their views. The conclusion of the Register Project means the end of this channel of communication – but the need for some such channel remains.

A number of persons have said that SCBDOK is unsuited to registers because certain parts of the template are only relevant for surveys. Therefore special instructions for registers should be drawn up. To simplify the work, existent manuals could be supplemented with examples of good documentations for both SCBDOK and Metadok. It is also essential that Statistics Sweden should offer a course on the metadata system as a whole and the use of the documentation tools. There used to be a recurrent course on SCBDOK but it was cancelled for lack of teachers. The documentations will not be of high quality unless there is training, not least because many persons find SCBDOK difficult.

> *21. There must be training and manuals concerning both the metadata system as a whole and its parts. We suggest that the network for register questions which is discussed in Section 10 should also function as a forum where views on the metadata system can be discussed and passed on. Furthermore we should like to draw particular attention to the need for special instructions to guide the persons documenting registers in SCBDOK.*

9.   Secrecy and protection of privacy

At the same as the register system is improved in order to facilitate joint processing and raise the quality of the statistics, secrecy and the protection of privacy within the system should be reinforced. It is essential that the preservation of secrecy be improved without delay.

Increased security for persons with protected identity

The fact that neo-Nazis obtained sensitive data from a person working at a social insurance office brings to the fore the question of how Statistics Sweden can improve its security routines with regard to persons with protected identity. There is a special variable in the Population Register indicating the need for protection of identity in each case.

These persons (of whom there are at present some 8,000) should be treated in one of the following ways:

– Information regarding address, parish and house property is removed from the Population Register and all other registers. The special variable is got rid of. Possibly the information can be stored in a file with limited access right.

– The persons are removed entirely from registers with civic registration numbers in the "ordinary" register system but are to be found in a special, well-protected register.

At present these persons are not included in our samples from the Population Register.

Increased protection of privacy

The debate concerning the 1990 population and housing census left a negative impression regarding Statistics Sweden, and the debate concerning the abortion register in autumn 2000 showed that this was still an extremely sensitive question. This brings up the question whether Statistics Sweden needs to improve its routines with regard to the protection of privacy before a new debate concerning the 2005 census arises.

Name, address etc en clair should only be in the basic registers. It is quite unnecessary that all Statistics Sweden staff who come into contact with other registers should be able to read such information. Even though this is not a major problem, all other registers must be made anonymous by way of numerical codes.

Can we reduce the use of civic registration numbers? The new legislation concerning statistics makes it possible to give out anonymised data furnished with serial numbers for subsequent updating. We might also use anonymised registers with serial numbers within Statistics Sweden. This presupposes that the keys linking civic registration numbers and serial numbers are well-protected. Serial numbers can be used for sensitive registers within Statistics Sweden which at present contain civic registration numbers or corporate registration numbers (which sometimes are civic registration numbers). If the 2005 population and housing census only makes use of serial numbers, Statistics Sweden will have the right to say that it has taken a major step in the protection of privacy.

In connection with this we should like to draw attention to the fact that that there has been expressed a desire that the rules for the giving out of microdata should be made more precise. Furthermore there is risk involved in the giving out of detailed metadata of a technical character.

22aIt should at the earliest opportunity be investigated how persons indicated as requiring special protection of identity can be accorded better protection within Statistics Sweden's register system. This goes not only for the Population Register but also for the Business Register and the Real Estate Register. Name, address etc. shall not appear in any of the system's registers except these three basic registers.

22b. Replacing civic registration numbers with serial numbers would constitute a major change, and there must be a careful investigation of the possible consequences.

22c. The present rules for the giving out of microdata should be made more precise: the attention of users and staff must be drawn to the fact the same rules apply to civic registration numbers in both business and individual registers. The rules must be made more concrete: Which variables are sensitive and which are less so?

10. Register function and network

It has for some years been a task of the Register Project to handle important questions of method and co-ordination in the register area. Now that the Project is coming to a close it is necessary to consider what further work in this area is required and what forms it should take. Therefore such questions should be more permanently anchored in Statistics Sweden's organisation. The same goes for the questions concerning metadata such as have been dealt with in project form during recent years. There are strong connections to central questions concerning standards, which are already form a part of the Methodology Unit's basic operations but for which the resources are few.

Proposals and initiatives

We have on a number of occasions taken up these questions in the Scientific Council and in conversations with the directorate. There has also been a proposal by Bo Sundgren (Government Official Reports 1999:96) that a register function should be established within the central database operations for the processing (including joint processing) of microdata for statistical purposes. The current investigation concerning a new methods organisation at Statistics Sweden points to the need for an enhancement of the central function for questions concerning standards. The new Metadata Project will have important tasks regarding documentation and the development of the metadata system.

Finally it may be noted that the R&D Department in its guideline document for 2001 has been assigned the task of investigating, together with the Unit for Publishing and the departments, what form a central register service function might take. Experience is available from Denmark and Norway, where the central offices of statistics have units for customer service, advanced register tasks and the passing on of microdata.

---

23. There should be a central register function within the R&E Department with responsibility for questions concerning development and co-ordination. To this function should be linked a network of staff from the rest of Statistics Sweden with methodological and production experience. The organisational questions taken up in this report should be dealt with in association with the R&D Department's current investigation regarding a central register service function.

Distribution of work

The new organisation at Statistics Sweden, dating from 1998, has involved changes in the placing of the basic registers and has brought about a better accord between the formal organisation and the register structure. Previously three of the basic registers were within one and the same department. In the present organisation the departments largely coincide with the four basic registers. This means that all departments are deeply involved in the register system, but also that the need for co-ordination has increased.

The Register Project's contact group has in recent years functioned as an important forum for the exchange of experience and the handling of practical questions of co-ordination. There is at present no permanent solution, and here a central function is insufficient. We propose a network to take over the role of the contact group. Furthermore working groups within this network can develop new types of joint processing.

Statistics Sweden needs to have the capacity to analyse large sets of microdata. It should be possible to develop this capacity within the frame of the network. Such a capacity is important for the development of new register statistics. Fur-

thermore researchers who work with data provided by Statistics Sweden may make errors of calculation or arrive at controversial results – then we at Statistics Sweden must be able to perform our own analyses.

The production of register statistics and the development of the register system call for co-operation across program and department borders. The distribution of roles between programs, departments and the central register function must be clear. In broad terms this distribution might have the following appearance:

24a. Every department is responsible for its basic register and other registers within its area of responsibility, and has special responsibility for a number of standardised variables.

24b. Every department is responsible for the keys between its statistical registers and the particular basic register.

24c. The development of register method questions should be carried out within Statistics Sweden's general methodology program.

We propose that the central register function should be located within the R&D Department and should have the following areas of responsibility:

25. The central register function is to be responsible for the following:
- *Overall questions of co-ordination and standardisation which have to do with the definitions and names of objects and variables.*
- *The metadata system and the development of documentation support.*
- *Seeing that general guidelines are followed.*
- *Maintenance of a network for the exchange of experience.*
- *Seeing that development of new joint processing and the analysis of competence occur within the frame of the network.*