FILES OF INDIVIDUAL DATA AND THEIR POTENTIALS FOR SOCIAL RESEARCH

BY ODD AUKRUST AND SVEIN NORDBOTTEN

Central Bureau of Statistics of Norway

In a statistical file system, data collected from different sources and at different dates are stored at an individual unit level. Thus, they may be linked and used for the preparation of statistics and for analytical purposes as the need arises. The statistical file system will provide a much improved data basis for empirical research in the social services including demography, economics, economic geography, sociology, education, labour, social medicine, criminology and social psychology.

I. THE BASIC IDEA OF THE STATISTICAL FILE SYSTEM[1]

The traditional way of producing statistics is to carry out a special survey or to take a census. In a statistical survey or census, data are usually collected by means of questionnaires answered by individual respondents. The data are processed once and for all and the results published as statistical tables. The individual data are of no practical value for future utilization because they are neither satisfactorily identified nor stored in an easily accessible form.

The statistical file system, in contrast to the traditional system, is based on the idea that individual data are systematized and stored in such a way that they can be retrieved and reprocessed when new needs appear. The technical opportunity for the statistical file system was generated by the introduction of the electronic data processing equipment which has made it possible to systematize, store, search, retrieve and link large volumes of individual data within a reasonable time and at acceptable costs.

An important condition for the establishment of the statistical file system is that a permanent and unique identifier is assigned to each respondent for which data are collected. These identifiers serve as "identification tags" on the data stored. In this paper, in which we are concened only with persons, personal identification numbers therefore play an important part in three aspects.

First, by means of the individual identifiers, it is possible to file the data collected in a systematic way. Second, the individual identifiers permit retrieval from the files of all data stored for a specified group of persons even though the data may have been collected from different sources and for different periods. The individual identifiers make it also possible to link the originally separated data in new and more application, oriented records for each inoividual, permitting us to carry out a wide range of cross-sectional as well as longitudinal studies [4].

The potential applications of a statistical file system depend of course on the data included in the system. In general—and in contrast to the traditional system—any additional data included do both contribute with their specific utilization value, but at the same time they also increase the utilization value of previously included data by increasing potential combinations of data.

[1] The present paper is based on article prepared in a popular form in Norwegian by the authors [1].

II. Needs for More Statistical Information

Central statistical institutes are confronted with an increasing demand for statistical information of four types:

(a) General statistics with standard regional specification.
(b) Special statistics with standard regional specification.
(c) General statistics with special regional specification.
(d) Special statistics with special regional specification.

To satisfy the general need and interest of the public for information about the country has been a well accepted task for the official statistical institutes.

An increasing interest for special "tailor-made" statistics to satisfy the individual user's need has recently been noted. Requests for special estimates and tabulations by central and local government agencies responsible for planning and development projects are frequent. Similar requests also come from the private sector which often wants information too special to be published in printed statistical reports or be included in a general tabulation. The users include also social researchers with problems not previously formulated in quantitative terms who may need special statistical information to test their models.

The special statistics demanded may range from simple frequencies to estimated structures of complex mathematical models. Very often data from different collecting operations are required for preparing these tailor-made statistics. The most efficient way to meet the demand seems to be by means of a statistical file system permitting a return to the individual data to carry out necessary record linkage generating the data base for the required statistical computations.

The demand for geographically specified statistics becomes more and more pronounced. Users, usually administrative managers, needing more statistical information for the standard administrative regions such as counties, municipalities, etc., can be distinguished from those needing information about otherwise delimited regions such as specified geographical polygons. The latter group is mainly represented by planners working on problems related to regional and urban development, communication, transportation network, pollution and environment.

The demand for regional statistical information may be met by introducing and maintaining a system of special units which can be linked to the units of the economic activity and population registers.

The foundation of demographic development and behaviour analysis is data about the population. Demographic studies have been an important field of research for hundred of years. However, there are still many questions to be statistically answered about fertility, nuptiality, migration, mortality, etc. It will be necessary to introduce a number of non-demographic factors into the analysis. Variables which can later be regarded as instrumental variables are of particularly great interest when the models are to be used for planning purposes.

The need for more knowledge about the propensities of the population to acquire education subject to different social and economic situations is frequently pointed out as an important element of the planned development. To be able

to stimulate the development of education in such a way that the natural background of the population is utilized in an efficient way, statistical knowledge is required about which factors affect the educational pattern and to which degree. Traditionally these factors have been studied by means of statistics on pupils by sex and age at different types of schools. Such information may be useful in preliminary research. However, estimating relations which can be used for planning and management purposes requires statistical information about the possible relationship between the demand for education and the personal and environmental factors, e.g. sex, age, previous education, marital status, income, educational level of the parents, their occupation, the family size and the general economic situation.

Another field for which planning activity requires more statistics is the labour market. So far, the most relevant statistics available have usually been the occupational frequencies by sex, age and region at different points of time. As a basis for the construction of planning models, these frequencies have a limited value. They do not, e.g., give much information about the probable effect on the labour market from changes in variables controlled by the government. More statistical information about the reaction of the labour force to alternative government actions is necessary if it is to be possible to check the labour force implications of economic plans. Education as well as previous professional experience seems to be among the central factors which must be linked to the occupational status.

Mortality statistics represent a well developed statistical area. We know pretty well the death cause frequencies by sex, age and region. Much less is usually known about the morbidity by illness, status and duration. In other words, we know much about the final outcome of illness, but little about the pains and problems caused by illness. In order to evaluate the efficiency of existing health services and develop rational plans for future extensions of the services, statistics on morbidity are demanded. The health services may in this context be regarded as a nation-wide process, the aim of which is to improve the health status of the population. The factors by which this aim can be attained may be the size and composition of medical personel, the type of medical institutions and their capacity, and the distribution between preventive and curative treatment. To plan the most efficient factor allocation for the process requires statistical knowledge about the relations between factors and results.

Morbidity is also an important aspect to be taken into account when the total labour potential of the country is to be analyzed. Health surveys which have been carried out indicate that a significant part of the input potential of the labour force is lost because of illness. To be able to attack the problem of how to increase the input of the labour force most efficiently by decreasing the morbidity will require further statistics on morbidity.

So far, studies on personal income distributions have mainly been limited to cross-sectional data on general income characteristics. It is obviously necessary to investigate the relation between personal income and consumption, education, occupation, industrial activity and taxation not only in a certain period but also over time to know what impact changes in these relations may have on the income distribution, and *vice versa*. An increasing interest in the

personal life income concept accentuates the need for statistics which show how cumulated income grows through the life cycle and why.

The high economic standard reached in the developed countries has changed much of the attention from economic toward non-economic development implications. Criminology, sociology, political behaviour, time budgets, social indicators, etc., are all fields of social research which require more comprehensive data which may help to explain quantitatively the consequences of alternative actions. Common to all the above mentioned problems is the need for statistical information which development of a statistical system of personal data files can give.

To ensure comparability and consistency, conceptual systems have to be established as a framework for the data registration. One task is to develop a socio-demographic accounting system based on a general set of classification systems. Another and even more fundamental task to is systematize in a catalogue of variables the individual characteristics which are being observed, recorded and stored and their definitions.

III. Data Collection and Storage

By a population register we mean a list kept on paper, punch cards, magnetic tape, etc., with name and address for all members of the population at a specified date. The statistical purposes of a population register are to form a frame for statistical investigation and to be a tool for integrating individual data. The population register must therefore be supported by a set of maintenance routines for including changes, increases and decreases. To be an integration tool, the population register must also include a permanent and unique identification symbol for each person [3].

A statistical file system requires collection of data identified to persons listed in the population register. The acquisition may be either direct statistical observation or transfer of data recorded as a part of an administrative process. The data collection may be a continuous observation of events or a periodic observation of the states of the population with respect to one or more characteristics.

The needs for statistical information outlined in the previous section will require substantially more data collection than most statistical systems collect today. Much of the additional data wanted are already recorded in some form through administrative processes carried out by government agencies. An important task is therefore to get these agencies to record the data in such a form that they can efficiently be included in the statistical file system. This may be possible by offering the administrative agencies register services in return, i.e. offering maintained lists with names, addresses and identification numbers for the population, but of course excluding all kinds of data which are considered confidential and personal.

Administratively recorded data will usually be complete in the sense that they are not samples. At least in small countries, the advantage of complete data compared with sample data seems to be large enough to justify complete storage. Data collected by direct statistical observations will often be on a sample basis

because of the significant observation cost. But even data from sample surveys should be identified by the identification numbers and included in the statistical file system.

The largest direct statistical collection of data regarding individuals is the population census. In a statistical file system, preaddressing and -identifying the census questionnaires by a population register is the obvious step towards an efficient identification of the census data. Among the questions on the census questionnaires is usually also one about education obtained. When these are appropriately formulated and repeated in two or more censuses, they can give a basis for statistics on the population flow pattern through the school system.

A more satisfactory collection of educational data would be transfer of recorded educational events from the school administration. If the individual applications, matriculations and graduations locally recorded anyway are identified by the population identification numbers, great advantages for both the school administration and the statistical file system may be expected. When the pupils first are admitted to the basic school, the identification numbers may be introduced in the school system by birth certificates issued with the numbers. Later, the identification numbers may follow the pupils by their graduation certificates.

Occupation and industrial activity are two other characteristics which have mainly been observed periodically by means of censuses. A continuous recording of status changes will, however, give a much more powerful statistical data base. The necessary reporting system will in most cases hardly be justified by statistical data needs only. But there are many aspects of the labour market which have to be recorded for administrative purposes in a modern, industrial country. The question is whether it is possible to coordinate all interests in a common system based on the population identification numbers.

An intermediate solution may be more frequent statistical censuses based on simple questionnaires with pre-printed identifications and addresses from the population register, and activity and occupational states from the last census. The attached instructions to the respondents may be to correct preprinted information only if any changes have occurred.

The area in which administrative interest for a common recording based on the population identification numbers is strongest at present seems to be in health administration. Also in this case direct periodic, statistical observations may be performed, but will then probably imply central identification of the medical data. In this area, however, a considerable amount of personal data are administratively recorded. An efficient recording seems to require two questions to be solved.

First, the doctors, hospitals and other health institutions have to get the identification numbers of their patients. One solution may be to make the patients acquainted with their numbers by issuing special "health cards" required to get medical treatment. Another more expensive solution would be identification of each patient by name, etc., in a look-up procedure using population register lists. Second, a considerable amount of standardization in the description of medical diagnosis, treatment and cost aspects of individual cases is needed.

A national health recording system will both increase the medical efficiency and the source for important epidemiological and socio-medical statistical research.

All the above discussed data collections within the requirements of a statistical file system may be performed according to two alternative approaches. The first is to record periodically the status of each person at specified dates by the use of pre-identified questionnaires. The advantage of this alternative is the independence of non-statistical agencies while the drawbacks are that all expenses have to be covered by statistical uses and that only the periodic status data can be obtained.

The second and more satisfactory alternative is to introduce the use of the population register identification numbers in as many as possible of the administrative processes recording individual events and acquire copies of the data for statistical purposes. The advantages are the exact time specific data on events obtained and the cost-sharing with administrative agencies. The problem is to get the different administrative branches to agree on the standardization of concepts and coordination required.

Storing the data in a statistical file system has both a logical and a physical aspect. The basic idea of the system implies that any data element must include: (1) identification of a person, (2) identification of the characteristic observed, (3) specification of the observation reference period or date, and (4) the observation result. Just as it is necessary to maintain registers of persons with cross reference between the external name and address and the internal identification of the person, it will also be necessary to maintain a catalogue of characteristics with cross-reference between the name and definition and the internal identification for the characteristics.

Fast access and linkage of data can be promoted by organizing the data in such a way that a minimum search for and sorting of data are required. In contrast to ordinary data bank retrieval, the retrieval from the files of a statistical system will be of data clusters. The stronger the intra-correlation cluster effect of the different applications the better will the possibility for an efficient search-minimizing organization of the data be. The linkage of data will mainly be by individual identification numbers. The obvious sort-minimizing ordering of data elements within a data cluster is therefore by identification number.

The physical storage will be either within a direct or a sequential access file. In the direct access file a given number of storage positions can be accessed in random order at the same speed, but so far it implies a high storage cost. The sequential access data file gives a much cheaper storage cost, but the storage positions are only available in a given sequence.

With the present cost relations the sequential access data file seems yet for some time to be the only reasonable choice for mass storage of individual data files of the size discussed here.

IV. Analytical Utilization of the Personal Data Files

A personal data file system, once established, will provide useful information for research in a variety of fields: demography, economics, education, labour

relations, social medicine, consumer behaviour, sociology, criminology, psychology, election research, economic geography, etc. In most cases the aim of this research will be to establish behaviouristic relationships with a view to their incorporation in socio-demographic models [3].

We shall assume in this section that a personal data file system has been established consisting of:

A general data file covering all residents, past and present, containing the following general background information on each individual:

(a) Demographic characteristics: Date of birth; marital status, past and present; number of children, living and dead; the year of birth of each child.

(b) Family characteristics: Identifier numbers of father, mother, spouse and children allowing e.g. (i) a study of a person's background as a child through information available about his parents, (ii) the identification and study of groups of persons related by blood.

(c) Place of residence, past and present.

(d) Educational characteristics, e.g., examinations passed, year of enrollment and of examination and (perhaps) marks obtained.

(e) Income and wealth characteristics (as assessed for taxation purposes).

(g) Economic activity, past and present, characterized by industry (an industry classification of the firm with which the individual is/was associated).

(h) Occupational characteristics, past and present.

A set of special data files, covering various sub-groups of the population and containing specialized information useful to the social sciences, e.g., electoral registers, hospital files, social security files, criminal records, etc., as well as *ad hoc* information collected through specialized statistical surveys (surveys of the farming population, surveys of individuals subject to social care, etc.).

We assume that the same system of personal identifiers has been used throughout, so that data contained in the various files can be easily linked both for transverse and longitudinal studies.

The potential usefulness of such a personal data file system for socio-demographic analysis would be very great indeed. In what follows we shall list, by way of examples, some fields of research which are likely to benefit from this type of data.

Demography

An important field of research is the construction of demographic models needed, among other things, for demographic forecasting. The construction of such models will involve:

(a) Studies of fertility as a function of i.a. age, marital status, number of children already born, occupation, place of residence, and income.

(b) Studies of mortality as a function of i.a. sex, age, industry, occupation, and place of residence.

(c) Studies of migration as a function of i.a. sex, age, family status, education, occupation, income, place of residence, and housing conditions.

Research in these fields requires information on demographic variables and on other background variables, much of which will be available in a personal data file (sex, age, marital status, income, etc.).

Economics

(a) Analysis of the distribution of income and wealth, by persons and/or by households, and study of factors which caused the observed distributions: Personal data files will allow simultaneous studies of variations in income and i.a. age, activity, occupation, place of residence, and education (perhaps even examination marks obtained). Selected groups of individuals (e.g. persons who left agriculture, town immigrants, persons who at some stage were subject to social care) could be followed through time with respect to their income and wealth structure.

(b) Taxation research, including the construction of models showing how the taxation system, or changes in it, affect different types of households. A large amount of the data requirements of such studies—demographic, economic and social characteristics of the taxpayer relevant for tax assessment—will be immediately met by a personal data file as outlined above.

Labour Relations

The construction of models needed for forecasting the labour supply by regions, activity and occupation, involving studies of:

(a) Frequencies of occupation by age, sex, marital status, educational background, place of residence.

(b) Choice of activity and occupation as a function of education, activity and occupation of parents, place of residence, earning possibilities, etc.

(c) Choice of place of work as a function of own place of birth (and/or place of birth of spouse?), location of schools attended, access to schools for growing children, etc.

(d) Recruitment of leaders as a function of education, past career, social background, etc.

Education

(a) Studies of the demand for education, by type of educational institution: Many of the variables to be treated as exogeneous in a set of demand equations for education will be available in the personal data files, e.g. the activity, occupation, income, educational background, and place of residence of parents having children at school-seeking age.

(b) Education as a stimulus to economic growth.

(i) One method of measuring the productivity of education is to study

income differentials resulting from different amounts of schooling. The personal data files will provide data for such studies.

(ii) The productivity of education may be studied alternatively by means of macro-production functions (for individual industries or by the total economy) treating "educational capital" as one factor of production amongst others. The personal data files will contain data which are essential for measuring the amount of educational capital available.

Social Medicine

(a) The frequencies of diseases by groups of persons with different background characteristics: Studies are feasible in which medical data drawn from some special file (e.g. hospital records) would be linked with background data available in the general files, allowing the life history of each patient to be characterized by age, activity, occupation, place of residence, etc.

(b) Analysis of mortality by cause of death amongst groups of persons with different social characteristics: Studies would be possible by linking information on causes of death reported on individual death certificates with background information on each person drawn from the personal data files.

(c) Diet and mortality: by linking data on consumers' expenditure patterns, drawn from household expenditure surveys, with data on causes of death reported in subsequent years for persons belonging to the households studied, possible relationships between consumption patterns and mortality may be verified.

(d) Problems of rehabilitation: the personal data files allow selected groups of individuals to be followed through time, thereby opening up possibilities for studying the rehabilitation of individuals who at some stage were subject to medical treatment. Any medically interesting group of individuals could be made the subject of such studies: Persons discharged from psychiatric institutions, handicapped children, a clientele of drug addicts, persons treated after traffic accidents, etc. The findings might throw light on the effectiveness of various forms of treatment.

(e) Data on diseases, if linked with information on parent-child relationships drawn from the personal data files, could conceivably be used for studying inheritance as a cause of diseases. However, this would require data on the diseases in question to be systematically recorded over a very long span of time.

Sociology

This is the one branch of the social sciences which, in the long run, perhaps stands to profit most from the establishment of a personal data file system. Recent literature abounds with studies based on empirical data, laboriously collected, which in the future may be extracted with comparative ease from the personal data files. Cases in point are:

(a) Studies of social mobility, that is, movements out of and into various social groups. Studies of this kind may be undertaken on a continuous basis to the extent that the groups focused on are defined by demographic and social characteristics recorded in the personal data files (age, sex, marital status, occupation, activity, income, wealth, family background, education, etc.).

(b) Studies of the size and changing composition of socially interesting groups, like "old people", "divorced women", "non-complete families", "unmarried mothers". One obvious technique might be to use information available in the personal data files for a general description of the groups in question, while supplementary information needed to illuminate special problems would be collected on a sample basis by interviews.

(c) Sociological studies based on the possibility of following selected groups of individuals through time: One may visualize studies in which the past history of a group of individuals, as described by demographic and social characteristics drawn from the personal data files, are used to explain present actions of the members of that group.

Criminology

(a) Types and frequencies of offences against the law in different social groups: Studies of the causes of crime could be attempted by linking data on violations of the law, drawn from criminal records, with general background data of the convicts (including their past history) drawn from the personal data files.

(b) Studies of recidivism: Groups of persons who were once in conflict with the law may be followed through time with a view to studying their later rehabilitation into society. Such studies might help to illuminate effects of various forms of legal reactions.

Social Psychology

The frequency of selected psychological reactions in different social groups: Studies are feasible which would link data collected for a special purpose by psychiatrists with general background data drawn from the personal data files.

Election Research

(a) The relative strength of political parties in various social groups: Studies are feasible which would link specially collected information on the party preference of individuals with data characterizing the social background of the individual in question drawn from the personal data file (age, occupation, income, family background, occupation of father, etc.).

(b) Studies of electoral participation: Observations of behaviour at elections drawn from electoral registers (e.g. voted/not voted) linked with relevant background data available in the personal data files.

Consumer Behaviour

(a) Marketing research: As part of a prediction of future markets, data on the present consumption pattern of different social groups may be combined with information derived from the personal data files on the size and probable future growth of the various groups.

(b) Marketing research for public services: What is the size and the composition of social groups to which different kinds of public services are being offered or contemplated?

V. Development in Norway

A statistical file system has been under preparation in the Central Bureau of Statistics of Norway since the early 1960's. The work on the personal data files was first concentrated on the task of establishing a complete and operational population register and promoting its use among administrative agencies.

A system with unique and permanent personal identification numbers was introduced in 1964. An identification number was assigned to each inhabitant covered by the 1960 Population Census and continuously maintained from 1964 as a part of the current population registration. The number comprises 11 digits, the six first of which refer to the date of birth while the following three digits

TABLE 1

Content and Maintenance Frequency of the Personal Data Files*

Individual Data on:	Maintenance Frequency		
	Current	Annual	Less Frequent
1. Date and place of birth	*(from 1964)		*(from 1960)
2. Residence	*(from 1964)		*(from 1960)
3. Name	*(from 1964)		*(from 1960)
4. Address	*(from 1964)		*(from 1960)
5. Marital status	*(from 1964)		*(from 1960)
6. Date and place of death	*(from 1964)		
7. Migration	*(from 1964)		
8. Mother's identification	*(from 1964)		
9. Father's identification	*(from 1964)		
10. Spouse's identification	*(from 1964)		
11. Income		*(from 1967)	
12. Property		*(from 1967)	
13. Taxes		*(from 1967)	
14. Education	*(from 1971)		*(from 1960)
15. Occupation			*(from 1960)
16. Industrial activity			*(from 1960)
17. Housing conditions			*(from 1960)
18. Election participation			*(from 1960)
19. Social support		*(from 1966)	
20. Criminal action		*(from 1966)	

* The years refer to the first period the respective individual data were identified by personal identification numbers.

distinguish persons born on the same date and at the same time specify sex. The two last digits are check digits computed according to the modulo 11 method.

The personal identification numbers are now used in almost all our statistical surveys and censuses of persons. Even more important is the fact that the identification number system has also been adopted by many administrative agencies from which the statistical bureau receives data. Among the most important are the population registration, the tax authorities, the national social insurance, the health administration, the school administration, etc.

The main individual data tagged with identification numbers are presented in Table 1. At present the content of the files increases by several million data records annually.

The personal data files have been used for a number of projects requiring individual record linkage. The most extensive linkage of individual records from the same source but from different periods is that which has been carried out regularly since 1964 on the vital data on births, marriages, deaths and migrations each time an updated population status file is required. Each observed event, of which there are about 1 million each year, has to be matched with the record for the corresponding person from the old file. Sometimes several status changes may appear for the same person and a chronological control is required. The linkage of the data records constructed from the 1965 Census of Fishermen with the records for the same individuals from the 1960 Population Census represented our first experience in linking data from two sources.

We have later done several very successful record linkage operations, the most interesting of which is perhaps the 1960 Population Census file record linkage with the population status files for the end of 1967 for analytical purposes to be described below. In connection with the 1970 Population Census, which was designed so as to satisfy linkage requirements, plans for linkage with the 1960 Population Census file as well as with the intermediate population status files have been specified. The linkage between the income record files and the census file is another obvious project.

The current observation of acquired additional education started up in 1971. Later the registration will be extended to individual observation of all persons starting in educational activities as well as those dropping out.

Several longitudinal studies have already been or are being carried out on basis of the data in the file system. The main application is the estimation of the structure of the demographic model [2]. The computerized version of this model re-estimates age-sex-regional specific probabilities for demographic events for each time the model is used.

Another example is the study of death probabilities during the period 1960–1967 by year of birth, sex, occupation, industry and education as observed in 1960, for the whole Norwegian population.

VI. Summary

The traditional way of preparing statistics is to collect data through a census or survey operation and to process these data once and for all into a set of tables. The introduction of electronic data processing machines has opened up the way

for the statistical file system. The basic idea behind this system is to identify the individual data collected by unique and permanent identification numbers assigned to each statistical unit, and to store the data in a systematic and easily accessible way. The gain from such a system is that data from different sources and times can be linked at an individual unit level and thus allow the preparation of statistics as the need arises, e.g. illuminating problems which were not foreseen when the data were originally collected.

The demand for statistical information is increasing fast. In addition to the need for general tabular information as published in the different series of statistical publications, there are needs for tailormade or special purpose statistical investigations and analyses which are being felt more and more by planners and administrators at different levels and in different geographical regions.

The data collected must be broadened both by direct observation and by careful compilation of data already recorded by non-statistical agencies for administrative purposes, keeping in mind all the time that the data must be identifiable by identification numbers.

A higher frequency collection of educational, occupational activity and health data represents some of the main characteristics of the registration system for individuals, the development of which one should aim at. It is essential that the relations to the units of other registers be observed.

It is suggested that the statistical file system will provide a foundation for a large amount of empirical research in the social sciences including demography, economics, economic geography, sociology, education, labour, social medicine, criminology and social psychology.

A system with a separate and permanent identification number for each individual of the population was introduced in Norway, October 1, 1964, as an extension of the current population registration. A considerable amount of identified individual data are now stored in the Norwegian file system. It has been used for several analytical studies which would not have been possible without the systematized individual data.

References

[1] Aukrust, O. and Nordbotten, S.: "Dataregistrering, dataarkiver og samfunnsforskning", *Artikler fra Statistisk Sentralbyrå*, Nr. 34, Oslo 1969.

[2] Gilje, E. and Nordbotten, S.: "A Demographic Model for the Norwegian Population and its Technical Characteristics", *Statistisk Tidskrift*, No. 1, Stockholm 1971.

[3] Nordbotten, S.: "Purposes, problems and ideas related to statistical file systems", *Bulletin of the International Statistical Institute*, Vol. 42, 2, Sydney 1969, pp. 733–750.

[4] Nordbotten, S.: "Individual data files and their utilization in socio-demographic model building in the Norwegian Central Bureau of Statistics", *Review of the International Statistical Institute*, Vol. 38, 2, The Hague 1970, pp. 193–201.