Nordisk Statistikermøde i København 11.-14. august 2010

The statistical archive system 1960-2010: A summary[1]

Svein Nordbotten[2], Professor Emeritus, University of Bergen, Norway

Abstract

Fifty years ago, electronic computers and shared administrative registers with permanent and unique object identifiers were discussed for integrating and storing micro data from several sources and time periods for extending the official statistical services. Today, a major part of the official statistical products in the Nordic countries are prepared directly or indirectly from data identified by public object identifiers. Effective use of these data has required new methods for collecting, transferring, editing and organizing data, and for estimation and quality evaluation of statistical products. The development has provided wider statistical coverage and saved the communities for unnecessary data collection expenses.

*Keywords: S*tatistical archive system, register-based statistics, statistical data warehouse, official identification system.

1. 1950-60s - An epoch of new ideas and possibilities

Two important events in the middle of the last century initiated the ideas to be discussed. The first was the availability of electronic computers, and the second was the concept of unique and permanent object identificators.

Punch card systems were used in statistical production since 1890, and made the pre-war required increase in production speed possible [Truesdell 1963]. However, this equipment lacked the desired flexibility for solving more complex logical and computational tasks.

The next technical revolution in statistical processing started with the introduction of the electronic computer *UNIVAC I* at the US Bureau of the Census in 1951 [US Bureau of the Census 2000]. It proved to be the kind of tool the statisticians needed, and other national statistical institutes, NSIs, followed the US Census Bureau and acquired electronic computers during the next decade. The computers were fast and had a storage capability which at that time represented a potential for a huge increase in processing speed compared with the punch card equipment. Because computers could be programmed, they also opened up for simulating intelligent human actions depending on

[1] This paper is a revised version of extracts from [Nordbotten 2010].

[2]Address: Rieber-Mohns veg 1, 5231 Paradis, Norway. Email: svein@nordbotten.com. Web site: http://www.nordbotten.com

Acknowledgement: The author thanks Statistics Norway for supporting the preparation of this paper. The author is solely responsible for opinions expressed in the paper.

the data being processed, and became soon an effective tool for controlling and correcting statistical micro data [Nordbotten 1963].

The idea of considering all the important micro objects of the society as interrelated and interacting elements of a system, as reflected already at the macro level by the *System of National Accounts* [Vanoli 2005], was intensively discussed in the 1950-1960s. The collection of data for parts of the system should not be considered as independent and carried out in separate divisions. The NSIs were conceived as *production systems* producing statistical *products* from different kinds of data frequently shared or reused in preparation of different products.

Early in the post-war period, the introduction of unique and permanent identifiers was considered as a means to more effective public administration. The Nordic NSIs became enthusiastic supporters of establishment and maintenance of central registers with unique and permanent object identifiers because such registers would increase the access to useful micro data and the possibility for organizing the data effectively by data linkage.

The paradigm for use of micro data with unique and permanent identifiers considered by the statisticians was based on 3 concepts:

1. Object identification
2. Time specification
3. Attribute observation

related in a 3 dimensional data space as indicated in Figure 1 in which each point referred to a data value. Planes or subsets of the space, each representing an interesting data subset, were recognized:

- The *object plane* representing all facts about a specific object,
- The *time plane* representing all facts for a given point of time,
- The *attribute plane* representing all facts for a certain attribute,

Along the object axis several types of objects, e.g. individuals, enterprises, and buildings, could be represented with a corresponding representation of their respective attributes on the attribute axis. The spatial object location was assumed to be an attribute value at a specified time. Ideally, the attribute value might be geographic coordinates or a link to an immobile object. If an attribute of object A referred to another object B of the same or another type, and the reference was by B's object identification, object A's attribute set could by linking be extended to comprise also all attributes of B. In this way, the description of one type of objects would not be limited to their own life-stories, but could be extended to other objects to which they were related. While navigating along the object axis was assumed to be done by means of object registers, a list of all attribute definitions, referred to as the *variable catalog*, was assumed for navigation along the attribute axis.

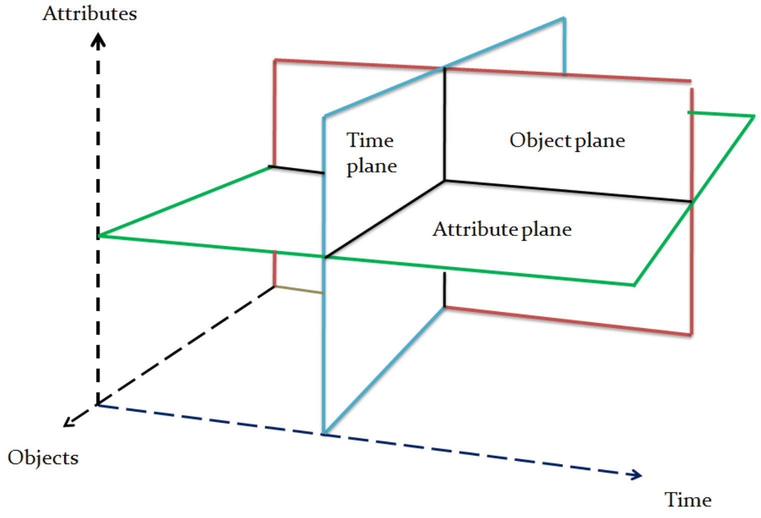

Figure 1: The data space

Data organized according to this model would by using the links to other objects permit latitudinal, longitudinal, spatial and relational studies which so far had been out of reach because the lack of permanent and unique identifiers made such linkage impossible.

Ideally, it should be possible to follow each kind of statistical micro object, such as an individual, enterprise, institution, commodity, etc. from one point of time to another, its interaction with other units and its movement from one location to another, i.e. it should be possible to follow and record any object from its *birth* to its *death* including its *relations* to other objects such as family members, employer, schools, and cars, and its *movement* in space over time [Nordbotten 1960/1967]. A more visual version of the model as shown in Figure 2 was frequently used [Nordbotten 1967]. It was presented as a storage container filled with small boxes of equal size. Each box was labeled with and ordered in the container by *object id*, *attribute id*, and *time id.* Within the box rested the observed data *value* of the fact.

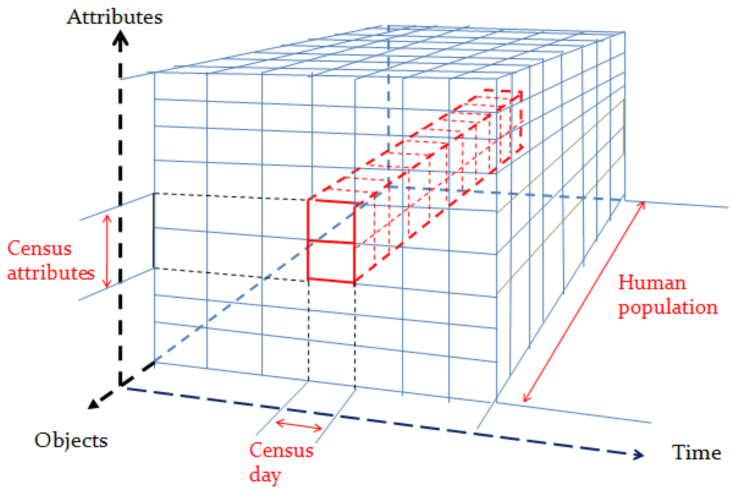

Figure 2: The data container as presented in the 1960s

The systematic data organization as illustrated by the data container also represented a basis for better *reuse of data* for different purposes compared with the separated data sets stored by source of collection.

In principle, as shown in Figure 3,, data for individual units could be retrieved from a wide network of interconnected data nodes and integrated to satisfy different application requirements.

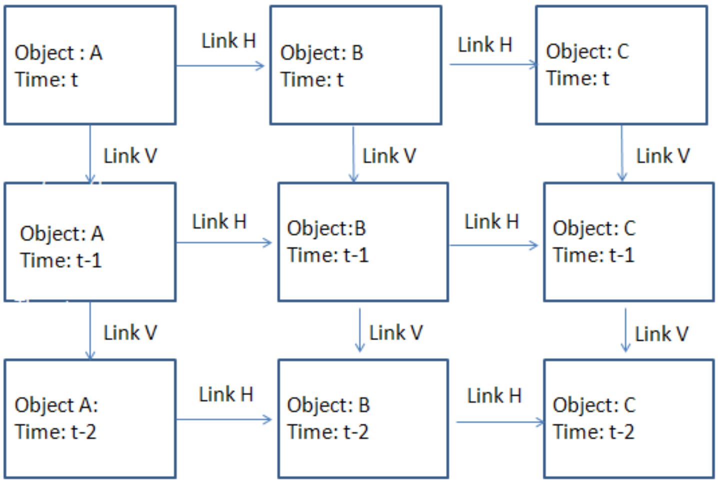

Figure 3: Integration of data sets by horizontal and vertical time linking

Data identified and organized in such a collection could by linking be used for producing a wide range of statistics, which had so far not been possible, and change the inflexible statistical production programs to a flexible on-demand production system. With permanent, unique and commonly used object identifiers established in central object registers, a huge mass of administrative micro data values could fill many boxes of the container. It was pointed out that since future use of collected data could not be anticipated, the data collection could be performed continuously and independently of a demand-driven processing and analysis.

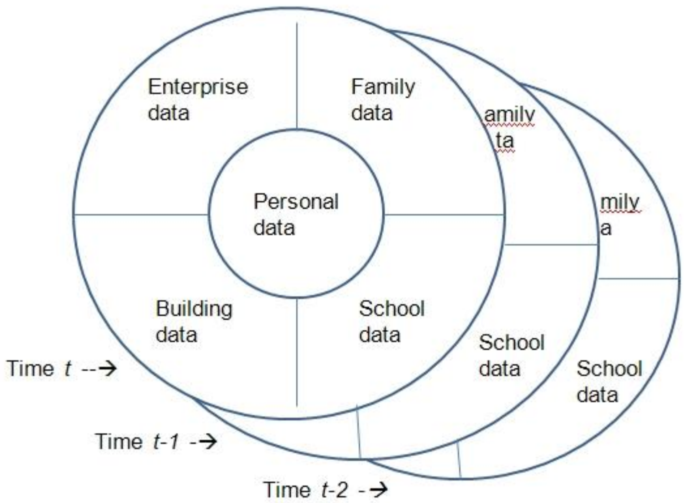

Figure 4: Integration of data sets by linking

As an example, Figure 4 illustrates how data for an individual could be integrated with data for other persons to whom she was related, to data of the enterprise at which she was employed, to data for the school she was attending and to data for buildings in which she lived and/or worked. Because of the permanence of the identifications, the integration could be extended by data for the same objects at previous points of time, i.e. a set of time series for a large set of attributes for the individual could be generated. The application value of an integrated set of attribute data will in general be far greater than the value of the same data considered separately.

Implementation of the organized data collection required establishment, use and maintenance of central registers for the major objects, i.e. for persons, enterprises and establishments, transportation objects, properties, buildings, etc. based on fast, powerful and reliable electronic computers. The presumption was that when operational, these registers would be used by public organizations saving resources by sharing the public expenditures of maintaining the registers.

In the Nordic countries, work on central registers started soon after the end of WW 2, and the initiative soon became popular also among private organizations such as banks and insurance companies. However, significant conceptual decisions had to be agreed upon as to definition of objects, frequency of updating and accessibility to mention only a few. The operational responsibility for each register had to be assigned to an existing or new organization referred to as register owners. Smooth working of these registers required willingness from register owners and users including the statisticians to compromise and adjust. Below we shall refer to the collection of these registers as the *official identification system*.

The *Statistical archive system* was introduced to the Helsinki Conference of Nordic Statisticians in 1960 as a data system in which all elements were directly or indirectly identified by identifiers from official identification systems [Nordbotten 1960/1967]. The visions were spelled out in a number of papers [Nordbotten 1968; Nordbotten 1970; Aukrust & Nordbotten 1970; Gilje and Nordbotten 1970]. Bo Sundgren presented in 1973 the theoretical foundation of the implicit statistical data bases in his doctoral thesis on the statistical data base system [Sundgren 1973]. The general ideas were discussed in a number of Nordic meetings and implementation projects were started in several countries. The most extensive project in the 1970s was the *ARKSY* project of Statistics Sweden.

In the envisioned statistical system, a conceptual distinction was made between *register* and *archive*. A register was a list of external descriptors necessary to locate each object, such as name, address, etc., associated with the unique and permanent identifier of the object and accessible to all users of the register. The external descriptors had to be continuously maintained, while the identifier should be time-invariant. The registers were interfaces between the external world and the data archive. The archive was, on the other hand, the collection of observations made of the objects' attributes, each in principle tagged with the respective object's identifier and time-stamp, and accessible only to the respective statistical authority. To support processing efficiency and confidentiality, the official identifiers might be converted to internal identifiers in the statistical archives. Other terms as *statistical file system* and *register-based statistical system* have been used as alternatives for the term *statistical archive system*. The concept *statistical data archive* has also been referred to as *statistical data warehouse.*

In the second part of the 1960s, it was proposed that registration of attributes in statistical registers of immobile objects such as properties and buildings should also include geographic location/coordinates contributed by agencies responsible for map development and maintenance. Population and enterprise objects could then indirectly be geographically characterized by the buildings and houses to which they were related at registration times. When implemented, this should permit computation of detailed migration, traveling, commuting and transport statistics, as well as development of required new special geographic classifications.

It was expected that IT and communication technology in the future would be generally available for government agencies, businesses and users in the private sector. Collection of data on electronic media was anticipated when the communication network capacity permitted. On the user side, it was assumed that the communication technology would also simplify the contact between statistical producers and users. The concept of *tele-photo-phone* was an imagined mobile device for the users. When the potentials became feasible technically, the demand for statistical information was believed to become more varied and special purpose to serve the different needs. Some kind of on-demand service had to be established to meet the demand.

The social scientists were expected to take advantage of the extended possibilities for longitudinal studies. In Norway, we established a new Socio-Demographic Research Unit within SSB to work with this kind of applications [Nordbotten 1970, Aukrust and Nordbotten 1970].

Technically, one serious restriction in the 1960s was the dominating serial storage media. Realization of many of the early ideas depended on availability to inexpensive and high capacity random access storage devices and efficient software for maintaining of and retrieval from data collections stored at such devices.

In the mid-1960s, a proposal for a *National Data Center* in USA based on motives similar to the Nordic work was set forward by the Ruggles committee of economists and social scientists [Ruggles, Miller, Kuh, Lebergott, Orcutt and Pechman 1965]. It ignited a heated data privacy debate and resulted in data privacy laws introduced in many countries during the 1970s [Miller 1967]. It was quite obvious that the legislation had to take into account the special task of the NSIs to maintain and manage the statistical knowledge about the society.

Figure 5 summarizes the main ideas and anticipations as to the statistical archive system in the 1960s.

Comprehensive official register systems
General utilization of systems

Fast and high capacity IT
Extensive and userfriendly communication
Software for data processing, saving, retrieval and communication

Data collection independent of specific usage
Use of communication network for collection
Functional organization

Wider statistical coverage by more intensive reuse of data
Statistics on demand
Analytical interest in and tools for exploiting the new possibilities

Legislation regulating the use of micro data

Figure 5: A summary of ideas from the 1960s

2. Development in technology and methodology

Few industries, if any, can present a faster technological development than the electronic computer and communication industries during the last 50 years. Online data storage capacity, processing speed and communication facilities have increased far more than anybody anticipated at the beginning of this period. From being an expensive and huge tool for a small group of mainly academic users, the IT-technology approaches today one billion users world-wide ranging in use from advanced research to everyday email and message exchange. The primitive concept of the tele-photo-phone has been replaced of handheld mobile devices for users' wireless communication with information providers including the NSIs.

Parallel to the technological development, statisticians have continuously refined their computer *processing* methods to take advantage of the new possibilities. Advances have been made in a number of fields from statistical systems and organization of the statistical production in general, to methods for on-line data collection and communication, data storage and retrieval, data editing and imputation, parameter estimation and predictions, on-demand access for users, and so on. Architectures for organizing computer facilities to suit the requirements of the individual NSI are available from single computers to advanced multi-computer clusters.

Some methods and implementations according to the 50 years old ideas and principles have been refined and become accepted also outside the Nordic countries [Houbiers 2004; Longva, Thomsen and Severeide 1998; UNECE 2007; Statistics Denmark 1995; Statistics Finland 2004; Wallgren and Wallgren 2007].

An important issue is how to organize the stored data for effective retrieval when requested. Since the 1960s, development of generic data base management systems have made significant progress from which the NSIs have benefitted. An outline of a structured data storage scheme for a process oriented statistical production was recently presented illustrating another step toward the systematic realization of the fifty years old dreams [Lundell 2009].

Descriptions and analysis of the implications of the statistical archive system ideas are presented in detail by authors who have actively participated in implementation and development of these ideas in Nordic statistics during the last 50 years [Gløersen and Gåsemyr 2010; Sundgren 2010b; Thygesen, L. 2010].

3. Applications in 2010

3.1 Types of applications in present statistical production

Official object identification numbers used in administrative applications, laws providing an NSI access to administrative data for statistical purposes, and technical possibilities for fast transfer of large data files, characterize environments in the Nordic and a few other countries making extensive use of data with permanent and unique object identifiers in their statistical work.

Use of this kind of data in production of statistics can be categorized by purpose as:

1. Controlling the processing and quality of micro data and evaluating final products,
2. Producing new statistical products either separately or in combination with data from multiple sources,
3. Preparing improved collection frames for sample surveys and censuses.

Examples of the first type applications are use of data integrated from several sources to support more effective editing and imputation of new data and assessment of the quality of population parameter estimates. The purpose of this type of use is mainly to improve quality of the statistical products. It provides also an extended basis for quality information about statistical products and information for future improvements of statistical processes [Granquist, Kovar and Nordbotten 2006].

The second category of application is continuation of the traditional utilization of administrative data when no adequate statistical data exist. The administrative data obtained from one source are now frequently linked to data from other sources thanks to the official register system. It provides a wider range of attributes for the objects and a potential for official statistics to provide more extensive and detailed description of demographic, social and economic aspects of the society. It saves the NSIs resources for data collection and provides a powerful basis for increased statistical quality. These possibilities have generated new methodological challenges for statisticians [Gåsemyr, Bjørke og Andersen 2007; Zhang and Nordbotten2008; Gåsemyr, Nordbotten and Andersen 2008].

Finally, administrative data have also proved to be very useful for adjusting and improving collection frames of statistical censuses and sample surveys by using available additional information about the objects.

Use of administrative compared with use of statistical data in a statistical application requires more attention since the administrative source has usually not tailored its collection according to statistical concepts, standards and requirements. Data from administrative sources may therefore need statistical pre-processing to solve intricate conceptual and matching problems before they can be used. The pre-processed data are typically organized by NSIs in *statistical base registers* for subsequent statistical processing.

3.2 Examples of present applications of statistical archive data

The greatest achievement in use of administrative data so far, is probably population censuses based completely on administrative data. The pioneering Danish 1981 Population Census based on stored administrative data was the first of its kind [Statistics Denmark 1995]. For some time, the housing part of the population censuses created problems for the NSIs because of difficulties in linking administrative data on buildings/homes/addresses to population data, but these problems seem now to be resolved.

Some examples of typical statistics based on administrative data today are:

- Census statistics (population and location registers)
- Population statistics (demographical registration)
- Foreign trade statistics(custom data)
- Income statistics (taxation data)
- Social statistics (registration of public services)

- Employment statistics (unemployment registration)
- Education statistic (registration of students)
- Health statistics (medical registration)
- Criminal statistics(judicial registration)
- Business statistics(enterprise registers)

As to economic statistics, general enterprise statistics directly from tax and account data are expanding, and we can see the possibilities for utilizing links between the different business actors as a first future step to economic infrastructure statistics. Linking enterprise objects to associated human objects utilizing their respective geographical attributes, offers another new dimension for economic statistics and analysis.

The extensive use of administrative data based on official identification systems has generated needs for new methods to handle problems created by this use. Examples of tasks requiring new or improved methods are:

- New electronic methods for effective transfer of administrative data from their sources to the NSIs
- Coverage control of administrative data before statistical use
- Improving sample survey frames and design taking advantage of administrative data
- Integration of administrative data from multiple sources
- Control and evaluation of the quality of integrated data sets
- Editing and imputing statistical data using background administrative data
- Estimating population parameters using background administrative data
- Privacy, confidentiality and security associated with statistical use of administrative data
- Quality metrics for and measurement of administrative data set quality
- Statistics on-demand from data archives
- Evaluation of cost and benefits using administrative data

The first task is dominated by legal considerations and negotiations concerning access to data from administrative authorities by the NSIs. When transfer procedures are established, the questions of integration of the administrative data with statistical data need attention. Linking data from different sources can present severe problems even if all components refer to the same identifier system. A set of administrative data can reflect more or fewer objects than expected by the statistical counterpart, time references can deviate, differences in definition of objects and attributes can exist, etc. This leads to the aspect of statistical quality of administrative data as perceived by the NSI, and how that quality should be evaluated. Even though a set of administrative data may be satisfactory for the purpose it was acquired, it may not be considered of satisfactory statistical quality. Effective statistical production based on linked files requires therefore new methods for editing, imputation, estimation and evaluation.

Measurement of the quality of administrative data sets is an important task also because an increasing future use must be expected to be on-demand, i.e. the users will demand statistical information when they need it, not years in advance. The NSIs should be prepared to respond with information on concepts, quality and cost of the demanded statistics based on data in their archives [Nordbotten 1993].

Many of these tasks are far from resolved, and there are still many lessons to be learned. Use of linked administrative micro data in research has made remarkable advances exemplified by a recent study of how mortality is influenced by the education of related people [Kravdal 2007]. The potentials of utilizing the longitudinal micro data are far from exhausted. More effective methods for access to, storage and use of data from administrative sources are needed.

In summary, countries with statistical production based on official identification register systems are at present saving valuable data from being wasted, their NSIs save resources using data already recorded, their respondents are spared for completing statistical requests about events they have already reported in administrative connections, and their official statistical products are increasing, but there are still potentials for improvements.

3.3 The personal privacy issue

The privacy problem as it emerged in the 1960s seems so far to have been resolved temporarily by the statistical laws and the introduction of privacy legislation as indicated in Figure 6.

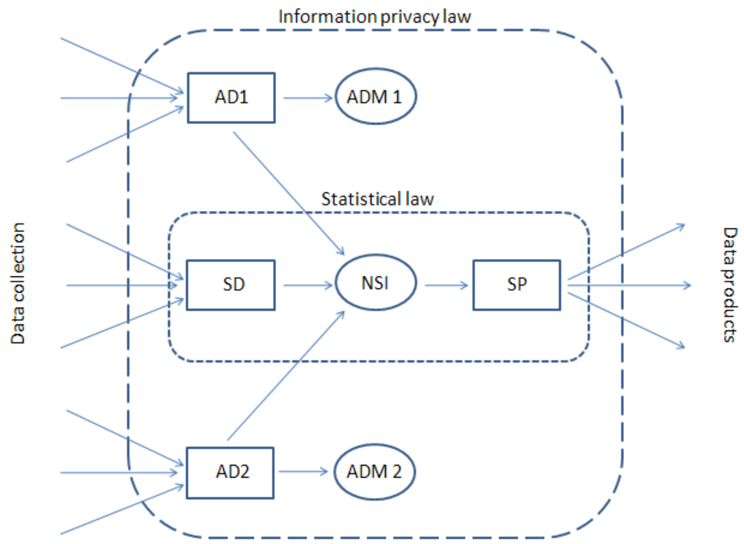

Figure 6: Legislation regulating the use by statistical and administrative agencies of personal data

On the collection side, the NSIs are allowed to use micro data collected for administrative purposes (AD1 and AD2) by other agencies (ADM1 and ASM2)in statistical production subject to the restrictions of the information privacy and statistical legislation. On the product side, certain dissatisfaction still exists with the access to micro data which will require future attention.

One latent problem is exposed by Figure 6. Is it possible to avoid a future user pressure to open the administrative data (AD1 and AD2) for other actors?

4. Concluding remarks

To conclude, a summary of the statistical archive system through 50 years may be done to compare the visions of 1960s with the reality of 2010. Returning to Figure 5, the answer may be illustrated by the following:

Comprehensive official register systems	Yes
General utilization of systems	Yes
Fast and high capacity IT	Yes
Extensive and userfriendly communication	Yes
Software for data processing, saving, retrieval and communication	Yes
Data collection independent of specific usage	?
Use of communication network for collection	Yes
Functional organization	Yes?
Wider statistical coverage by more intensive reuse of data	Yes
Statistics on demand	?
Analytical interest in and tools for exploiting the new possibilities	Yes?
Legislation regulating the use of micro data	Yes

Figure 7: Anticipations/realization through 50 years

Official register systems have been developed and accepted. The population register systems have probably also been more accepted and used by private organizations than could be dreamed of 50 years ago. The register systems for economic units are well established and their importance will probably be growing in the years to come, while the registers for geographic location seem to be working and will probably be improving in the future.

The technical conditions for realization of statistical archive systems are far beyond expectations. Nobody anticipated in the 1960s the present speed and capacity available in the current information technology or today's extensive mobile communication technology making any register object in principle available online.

Generic software tools for data collecting, processing, storage and retrieval have been developed and are available. The potentials of the hard and soft technology are far from exhausted and will probably open up for further improvements of the statistical archive systems.

The technology is used for collecting statistical as well as administrative data by means of the communication networks. However, the idea of making the data collection floating completely independent of the applications is still an alternative for the future, but there seems to be a trend toward organizing the collection tasks in independent units separated from the subject matter units.

The introduction of the public identification systems has increased the coverage of official statistics enormously compared with the situation 50 years ago when linking of data were almost non-existent, and by means of Internet connections, statistics have become a real public service. However, the demand for special statistics has not grown as fast as could be expected, and the capacity to provide statistics on-demand, i.e. immediate response to online queries for specialized statistics is far from being developed.

Fifty years ago, it was expected that the statistical archive system would create a heaven for social scientists. The use has, however, been restricted by varying views as to researchers' access to micro data.

What has the next 50 years in store for the statistical archive system? It was hinted to above that the technology already permit that statistical objects to be connected wirelessly to data collecting systems either in special situations, periodically or permanently. This can result in huge streams of micro data. To which extent these data could and should be utilized by official statistics is a political question, but most probably the answer will imply great challenges to the NSIs.

1. References[3]

Aukrust, O. and Nordbotten, S. (1970): *Files of Individual Data and their Potentials for Social Research*. Review of Income and Wealth.

Gilje, E. and Nordbotten, S. (1970): A *Demographic Model for the Norwegian Population and its Technical Characteristics*. Statistisk Tidskrift. No 2, Stockholm.

Granquist, L., Kovar, J. and Nordbotten, S. (2006): *Improving Surveys – Where Does Editing Fit In?* Chapter 4 in Statistical Data Editing, Vol.3: Impact on Data Quality. UNECE, Geneva 2006.

Gåsemyr, S., Børke, S. and Andersen, M.Q. (2007a): *A Strategy to Increase the Use of Administrative Data within an Integrated System of Business Statistics*. Seminar on registers in statistics - Methodology and quality. Helsinki.

Gåsemyr, S., Nordbotten, S. and Andersen, M.,Q.(2008a): *Role of Editing and Imputation of Sources for Structural Business Statistics*. Proceedings from UN/ECE Workshop Session on Statistical Editing in Vienna. Geneva 2008.

Gløersen, R. and Gåsemyr, S.: *The significance of IT development to archive statistical solutions.* Nordisk Statistikermøde i København 2010.

Hoem, J. and Nordbotten, S. (1969): *Some Model Types for Labor Force Projections*. Proceedings from 34. Session of International Statistical Institute. Contributed papers. pp. 29-39.

Houbiers, M. (2004): *Towards a Social Statistical Database and unified estimates at Statistics Netherlands*. Journal of Official Statistics, Volume 20, No. 1.
IRIDIUM (2009): http://www.cybergraphy.com/English/gps_satellite.htm

Kravdal, Ø. (2007): *A Broader Perspective on Education and Mortality: Are We Influenced by Other People's Education?* Social Science and Medicine. doi:10.1016/socscimed.2007.10.009.

Koren, J. (1914*): The History of Statistics, Their Development in Many Countries*. Book I: Historical and Commemorative Addresses. American Statistical Association. Macmillan. NY.

Longva, S., Thomsen I. and Severeide P.I. (1998): *Reducing Costs of Censuses in Norway through Use of Administrative Registers*. International Statistical Review, Vol. 66.

Lundell, L.-G. (2009): *Strukturerade datalager för effektivare produktion*. Unpublished paper. Statistiska Centralbyrån, U/ARK. Stockholm 2009.

Miller, A. (1967): *The National Data Center*. The Atlantic. Reproduced at http://blog.modernmechanix.com/2008/03/31/the-national-data-center-and-personal-privacy/

[3] Full-text copies of publications by the author are available at: http://nordbotten.com/articles/

Nordbotten, S. (1960/1967): *Elektronmaskinene og Statistikkens Utforming i Årene Framover* (*Computers and the Future Form of of Statistics*). De Nordiska Statistikermøter i Helsingfors 1960. Helsinki. pp. 135-141. Translated to English and published as: *A Statistical File System*. Statistisk Tidskrift 1966:2.

Nordbotten, S. (1967): *Purposes, Problems and Ideas Related to Statistical File Systems*. Proceedings from the 36. Session of the International Statistical Institute. Sydney 1967.

Nordbotten, S. (1968): *Automatic Files in Statistical Systems*. Statistical Standards and Studies, Handbook No. 9, United Nations, N.Y. 1967.

Nordbotten, S. (1970): *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, No.2. The Hague 1970.

Nordbotten, S. (1993): *Statistical Meta-Knowledge and –Data*. Statistical Journal of UN/ECE, Vol. 10, No.2, Geneva 1993, pp. 101-111.

Nordbotten, S. (2010):*The Use of Administrative Data in Official Statistics – Past, Present, and Future – With Special Reference to the Nordic Countries*, Official Statistics – Methodology and Applications in Honour of Daniel Thorburn, pp. 203-223. Available at Officialstatistics.wordpress.com.

Ruggles, R., Miller, R., Kuh, E., Lebergott, S., Orcutt, G. and Pechman, J. (1965): *Report of the Committee on Preservation and Use of Economic Data*. Available online at http://www.archive.org/details/ReportOfTheCommitteeOnThePreservationAndUseOfEconomicData1965

Statistics Denmark (1995): *Statistics on Persons in Denmark – a Register-based Statistical System*. Copenhagen.

Statistics Finland (2004): *Use of Registers and Administrative Data Sources for Statistical Purposes – Best practices in Statistics Finland*. Helsinki.

Sundgren, B. (1973): *An Infological Approach to Data Bases*. Statistiska Centralbyrån. Stockholm.

Sundgren, B. (2004): *Statistical systems: some fundamentals*. Statistiska Centralbyrån, Stockholm. Available online at http://sites.google.com/site/bosundgrenknowledgebase/

Sundgren, B. (2010a): *The System Approach to Official Statistics*. Official Statistics – Methodology and Applications in Honour of Daniel Thorburn, pp. 225-260. Available at Officialstatistics.wordpress.com.

Sundgren, Bo (2010b): *Statistical File Systems and Archive Statistics – Experiences from Statistics Sweden*. Nordisk Statistikermøde i København 2010.

Truesdell, L.E. (1963): *The Development of Punch Card Tabulation in the Bureau of the Census 1890-1940.* US Department of Commerce. Bureau of the Census. Washington, DC.

Thygesen, L. (2010): *The importance of the archive statistical idea for the development of social statistics and populations and housing censuses in Denmark*. Nordisk Statistikermøde i København 2010

UNECE (2007): *Register-based statistics in the Nordic Countries*. Review of best practices with focus on population and social statistics. United Nations Economic Commission for Europe. Also available online at http://www.unece.org/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

US Bureau of the Census (2008): *Factfinder – History and Organization*. Department of Commerce. Washington, DC. Also available online at http://www.census.gov/prod/2000pubs/cff-4.pdf

Vanoli, A. (2005): *History of National Accounting*. IOS Press.

Wallgren, A. and Wallgren, B. (2007): *Register-based Statistics: Administrative Data for Statistical Purposes.* John Wiley & Sons Inc.

Zhang, L.-C. and Nordbotten, S, (2008): *Prediction and Imputation in ISEE – Tools for More Efficient Use of Combined Data Sources*. Proceedings from UN/ECE Workshop Session on Statistical Editing in Vienna. Geneva.