Official Statistics in Honour of Daniel Thorburn, pp. 205–223

The Use of Administrative Data in Official Statistics – Past, Present, and Future – With Special Reference to the Nordic Countries

Svein Nordbotten[1]

The rolls, from which conscripts were called for working on the pyramids, may have been among the first administrative data sources for preparing statistics. Later rulers enrolled males for recruiting members to their legions. The enrolment data were probably also frequently used for statistics to assess military power. Since National Statistical Institutes were founded, many kinds of administrative data have been used. Electronic computers and shared administrative registers with permanent and unique object identifiers as applied in the Nordic countries, made it possible to integrate micro data from several sources and time periods, extending the set of possible questions to which statistics could provide answers. Today, a major part of the official statistical products are prepared directly or indirectly from administrative data. Effective use of these data has required the development of new methods for the transferring and editing of data, as well as for the estimation and quality evaluation of statistical products. Large amounts of data are now recorded automatically in connection with credit card use, electronic tickets, traffic surveillance and radio frequency identification of commodities, etc. and may be potential future sources of administrative data for official statistics. Technology trends indicate that embedded and implanted electronic chips will also be generating huge masses of administrative data. The threats to privacy are obvious, and possible use of such in the production of official statistics requires careful consideration, discussion and, if use is decided upon, careful preparation and implementation.

Key words: Administrative data sources; statistical archive system; register-based statistics; official identification systems.

1. Administrative and Statistical Data

In this paper, the distinction between administrative and statistical data for producing official statistics is rooted in the primary purpose of the acquisition of data:

1. *Administrative data* are collected primarily for nonstatistical purposes, and adopted for producing statistics,
2. S*tatistical data* are collected for the purpose of preparing statistics and are in general not available for any other purpose.

Typically, administrative data have been summarized for centuries by more or less advanced statistical processes to provide statistical overviews suitable for evaluation and

[1] Professor Emeritus, University of Bergen, Norway. Rieber-Mohns veg 1, 5231 Paradis, Norway. Email: svein@nordbotten.com. Web: www.nordbotten.com
Acknowledgments: This chapter is based on a keynote speech at the 2008 Conference of the International Association of Official Statistics, October 14, 2008 in Shanghai. The author thanks Statistics Norway for supporting the preparation for this speech and the further work on the present chapter.

decision making. Statistical data are specified and collected for the sole purpose of describing the status or development of a specific population when no other adequate data are available, and are seldom permitted used for other purposes.

Official statistics indicates that the producer is a *National Statistical Institute (NSI)*. One description of official statistics is: "*Almost every country in the world has one or more government agencies (usually national institutes) that supply decision-makers and other users including the general public and the research community with a continuing flow of information . . . . This bulk of data is usually called official statistics. Official statistics should be objective and easily accessible and produced on a continuing basis so that measurement of change is possible*" (Biemer and Lyberg 2003).

2. The Past

2.1. Early Use of Administrative Data

It is difficult to state when official statistics were first produced and used, but it was probably when rulers of communities wanted to compare their power with that of their enemies. Males have most likely been enumerated and records summarized to provide the rulers with needed information on which strategic decisions could be based. According to a Canadian source, enumeration of different resources was already being regularly carried out in Babylon (Statistics Canada 2009).

About 4–5 thousand years ago, Egyptian pharaohs also carried out censuses for tax gathering and to determine fitness for military and labor services as well as for surveying construction progress.

It became early quite common in many countries to enumerate the male population within certain age brackets in order to provide the rulers with statistics about their potential power, and records for recruiting soldiers to their legions. These statistics have in more recent times been utilized retrospectively to estimate the total populations and their age distributions.

Recording land properties was a usual means for rulers to determine taxation of their populations. Based on statistics from these records, heads of countries could evaluate their potential income and wealth as well as keep control by means of the collection process. Trade in commodities passing frontiers was another early source for collecting taxes, and the first international trade statistics appeared (Figure 1).

The first formal offices for official statistics were established in the 18th century. They were frequently named *Table offices*, reflecting the fact that their purpose was to summarize administrative micro data into *tables* of macro data, not to collect the data themselves (Koren 1914). At this time, demographic records, tax data, public accounting data, health data, social data, medical data and school data were aggregated to separate types of statistics to describe the prosperity of the country.

In the middle of the 19th century, international cooperation on official statistics was initiated. Some countries established *National Statistical Bureaus*, responsible for all official statistics, while other countries chose to organize statistical departments within several ministries. In both cases, the statistics prepared were mainly based on administrative data, later to be supplemented by data collected solely for statistical

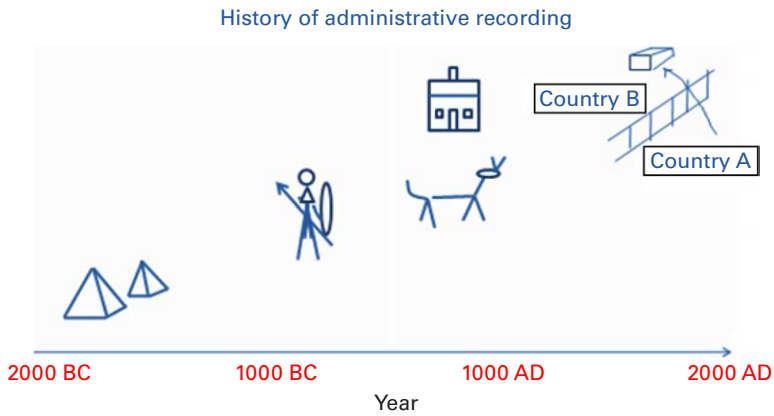

Fig. 1. Statistical use of administrative data

purposes such as population censuses and statistical sample surveys. Typical for the official statistics up to the Second World War was that the collection of data and production of statistics on different matters was to a large extent carried out independently.[2] This made integration of the statistical results rather difficult.

After the Second World War, the need for creating comprehensive and consistent descriptions of the economic, demographic and social aspects of countries increased, and particularly the *National Accounts System* became an important vehicle for organizing economic statistics into a conceptually consistent system (Vanoli 2005). However, because of the diversified nature of the data on which the different parts of the national accounts were prepared, the compilation of national accounts became a very complex operation.

The intention of developing a similar *System of Social and Demographic Accounts* was never realized.

2.2. *The Post-war Period of Great Ideas and Possibilities*[3]

Punch card systems were used in statistical production as of 1890, and they made the prewar required increase in production speed possible (Truesdell 1963). However, this equipment lacked the desired flexibility for solving more complex logical and computational tasks.

The real revolution in statistical processing started with the introduction of the electronic computer *UNIVAC I* at the U.S. Bureau of the Census in 1951. It proved to be the kind of tool the statisticians needed, and other national statistical institutes soon followed the U.S. Census Bureau and acquired electronic computers. The computers were fast and had a storage capability which at that time represented a potential for a huge increase in processing speed as compared with the punch card equipment.

[2] In his works on statistical systems, Bo Sundgren has named this type of statistical system "stovepipe architecture," and points out that remains of it can still be observed in NSIs (Sundgren 2004; Sundgren 2010).
[3] This section reflects the discussions in the Nordic countries.

Because computers could be programmed, they also opened the door to conditional processing, which permitted simulating intelligent actions depending on the data being processed or the type of application, and the computer soon became 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 on the macro level by the *System of National Accounts,* emerged and was intensively discussed in the 1950s–1960s. The collection of data for the system should no longer be considered as independent and carried out in separate compartments. The NSIs were conceived as *production systems* producing statistical *products* from different kinds of data frequently shared or reused in the 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 the 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 of organizing the data effectively.

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

1. Object
2. Time
3. Attribute

related in a 3-dimensional data space in which each point referred to a data value.

The spatial object location was assumed to be an attribute value at a specified time. Ideally, the corresponding attribute value might be geographical coordinates. Planes or subsets of the space were recognized as each representing an interesting data subset:

- The *object plane* representing all the facts about a specific object,
- The *time plane* representing all the facts for a given point of time,
- The *attribute plane* representing all the 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. 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 also include all the 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 (Figure 2).

Data organized according to this model would through using the links to other objects permit longitudinal, spatial and relational studies which so far had been out of reach because the lack of permanent and unique identifiers used by all data-collecting authorities made such organization impossible.

Ideally, it should be possible to follow each kind of statistical micro object, such as an individual, enterprise, institution, commodity, or service, from one point in 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* as well as its

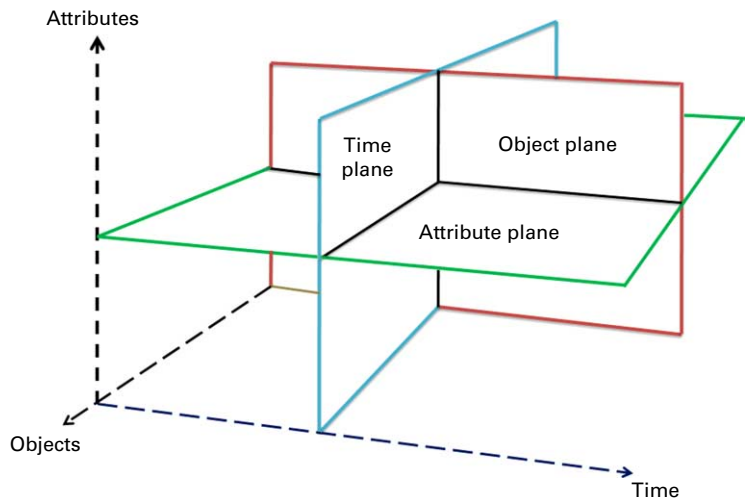

Fig. 2. The data space

relations to other objects such as family members, employer, schools and cars, and its *movement* in space over time (Nordbotten 1960/1967).

As an example, Figure 3 illustrates how data for a person could be integrated with data for other persons to whom he/she was related, to data of the enterprise at which he/she was employed, to data for the school he/she was attending and to data for buildings in which he/she lived and/or worked. Because of the permanence of the identifications, the integration could be extended by data for the same person at previous points in time, i.e., a set of time series for a large set of attributes. An important argument in this connection was that the application value of an integrated set of attribute data in general would be greater than the value of the same data considered separately.

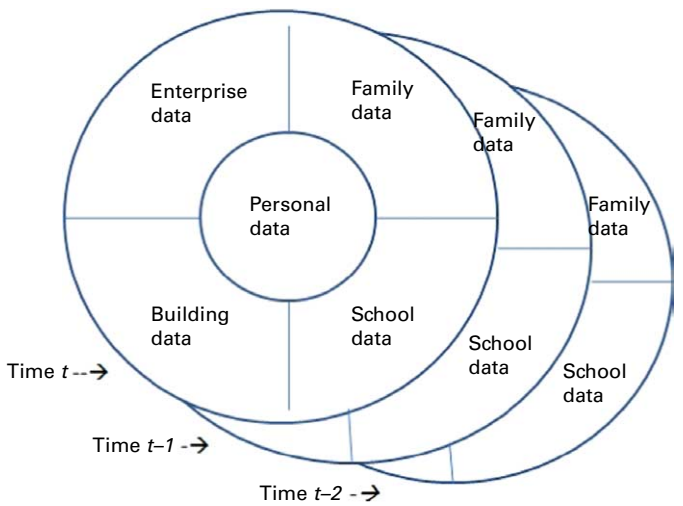

Fig. 3. Integration of data sets by linking

A more visual version of the model as shown in Figure 4 was frequently used (Nordbotten 1967). It was presented as a storage container filled with small boxes of equal size. Each box was labelled with and ordered in the container by three attributes: *object id*, *attribute id*, *time id*, and within the box itself rested the observed data *value* of the fact.

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 changing the inflexible statistical production programmes to a very flexible on-demand production system. With permanent, unique and commonly applied 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 the collection of data could be performed continuously and independently of the demand-driven processing and analysis.

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 costs of maintaining the registers.

In the Nordic countries, work on central registers started in the early 1950s, and the initiative soon became popular also in 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 the register owner. Smooth working of these registers required willingness on the part of the register owners, users including the statisticians, to compromise and adjust. Below we shall refer to the collection of these registers as the *official identification system*.

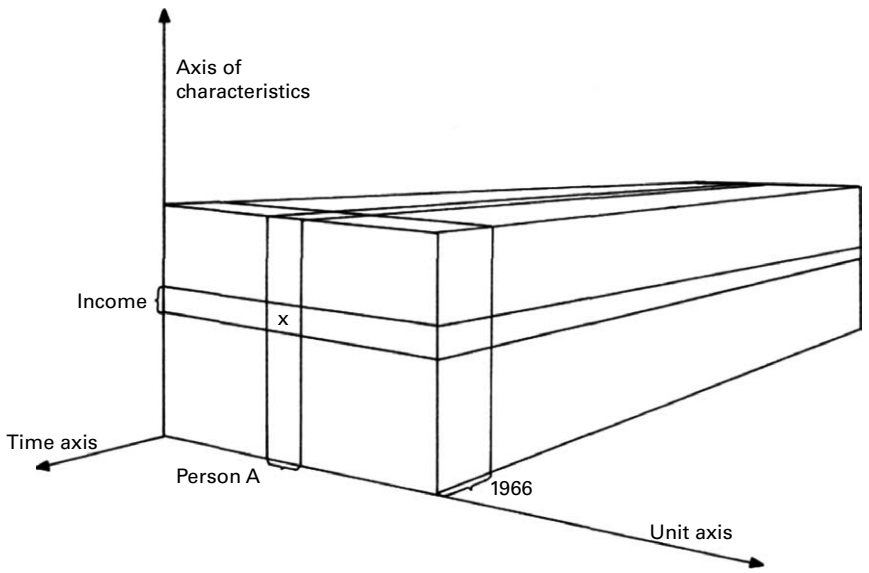

Fig. 4. The data box as presented in the 1960s

The *Statistical archive system* was introduced at 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 and Nordbotten 1970; Gilje and Nordbotten 1970). In 1973 Bo Sundgren presented 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 identifiers necessary to locate each object, such as name, address, etc., associated with the unique and permanent identifier of the object. The external identifiers 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 the collection of observations made of the objects' attributes, each in principle tagged with the corresponding object's identifier and time-stamp. To support processing efficiency and confidentiality, the official external identifiers might be converted to internal identifiers in the statistical archives. Later, other terms such as *register-based statistical system* have replaced the term statistical archive system, hiding the distinction between the public register system and the internal data systems.

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

In the mid-1960s, a proposal for a *National Data Center* in the U.S.A. based on motives similar to those behind 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 the 1970s in many countries (Miller 1967). The duties and responsibilities of the NSIs varied with the statistical and data privacy laws in the respective countries. In some countries the statistical agencies got exclusive access to administrative data while in others they did not.

3. Present Use of Administrative Data

3.1. 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, IT-technology approaches today one billion users, ranging in application from advanced research to everyday email and message exchange.

Parallel to the technological development, statisticians have continuously refined their computer *processing* methods to take advantage of these new possibilities. Advances have been made in a number of fields, from 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 for the use of administrative data in statistical production in accordance with the 50 years old ideas and principles have been refined and become accepted as *Register-based statistics*, 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, the development of modern data base management systems has made significant progress from which the NSIs have benefitted. Recently, an outline of a structured data storage scheme for a process-oriented statistical production was presented, indicating another step towards the realization of the fifty-year old dreams (Lundell 2009).

3.2. Types of Administrative Data Applications in Present Statistical Production

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

The use of this kind of administrative data in the production of statistics can be categorized by purpose as:

1. Controlling the processing of statistical data and quality evaluation of 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 of applications are the use of administrative data to support more effective editing and imputation of statistical data and quality assessment of population parameter estimates. The purpose of this type of use is mainly to improve quality of the statistical products. It also provides 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 the 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 the 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, and Andersen 2007; Zhang and Nordbotten 2008; 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 not usually tailored its collection in accordance with statistical concepts, standards and requirements. Data from administrative sources may therefore need statistical preprocessing to solve intricate conceptual and matching problems before they can be used. The preprocessed data are typically organized by NSIs in *statistical base registers* for subsequent statistical processing.

3.3. Examples of Present Applications of Administrative Data

The greatest achievement in the use of administrative data so far, is probably population censuses based completely on administrative data. The Danish 1980 Population Census was the first of its kind (Statistics Denmark 1995). For some time, the housing part of the census 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 (customs data)
- Income statistics (taxation data)
- Social statistics (registration of public services)
- Employment statistics (unemployment registration)
- Education statistics (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 of 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:

- Effective electronic 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 question of integration of the administrative data with statistical data needs 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 the 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 for which it was acquired, it may not be considered of satisfactory statistical quality. Effective statistical production based on linked files therefore requires 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 the quality and cost of the requested statistics based on data in their archives.

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 as 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 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 the task of completing statistical requests about events they have already reported in administrative connections, and their official statistics are increasing, but there is still potential for improvement.

4. Future Use of Administrative Data

4.1. Use of Administrative Data in the Next Decade

So far, most of the administrative data acquired by NSIs for producing statistics have been data collected by central, government agencies. But there are an increasing number of public as well as private organizations, e.g., special public services, banks and insurance companies, collecting data based on the official identification system to perform their respective tasks. A central question will be if these data are wanted by and can be made available to the NSIs for producing official statistics. The answer will depend on political decisions about social benefits of statistical information and risk of misuse of sensitive data.

The Nordic NSIs are in a unique position to utilize administrative data from new sources since their official identification systems are extensively used and the NSIs already have authorized access to produce statistics from multiple administrative data sources using the official identification systems. Other organizations than the NSIs will probably request similar access, but the probability of them being provided with similar authorization is small, this because of confidentiality concerns and the NSIs' high performance.

The crucial dilemma is whether information embedded in the data collected for administrative purposes should be wasted because of the potential risk of misuse or made available to NSIs for producing official statistics to increase the understanding of the complexity of our societies and contribute to higher efficiency. Extending access to new administrative data by trusted NSIs seems to represent a smaller risk of misuse than authorizing one or more other organizations to access such data.

Which, then, are the existing data sources and which are the statistical products of interest? A number of data sources already exist, which have in common that the objects they observe are identified within the official identification system and the attribute values observed are electronically recorded. These data are frequently referred to as *electronic tracks*, *electronic footprints*, etc.

A few examples of such data are:

- Data from public electronic services
- Credit card transactions data
- Tracking data by RFID technology
- Travel data from electronic ticket records
- International migration data from electronic passports

A number of public services are currently offered on the Internet, for example registration of new addresses, income declarations, social benefits, medical services, etc. They all have in common that they save the users the time needed for visiting the service organizations in person and save the respective administrative organizations the unnecessary use of resources. Statistics on who are the users of these services will provide information on the use of the services, the time and cost savings compared with traditional service deliveries, the possibilities of service improvements, and so on.

An increasing amount of commodities and services are paid for by means of credit cards. Since credit cards in the Nordic countries are frequently connected to accounts

identified by the owner's official identifiers, the data recorded in connection with purchases and services paid for by credit cards can, when linked to other data about the users, provide a basis for more detailed statistics about private consumption and expenditures, e.g., which type of people are making the purchases in which type of shops, and at which time, how far away from their registered home address, and changes from previous periods for groups of objects.

In the last 5–10 years, a technology known as *Radio Frequency Identification (RFID)* by which object identification is stored in small wireless chips embedded in objects that can be read by readers near the object has been expanding (Wikipedia 2009a; Miles, Sarma, and Williams 2008). The RFID chips can either be *passive*, i.e., obtain their operating energy wireless from the reader/environment, or *active* with their own energy sources. This technology is being increasingly used in many fields, for instance as a substitute for bar codes because the RFID does not require a nearby aimed optical reading device and because of low prices (about US$ 0.10 per chip). The RFID can be used for direct identification of commodities in a basket by a reader at the cash-register, and the amount to be paid instantly computed. An electronic time-stamped record with specification of the commodities in the basket with their prices will exist and can, if the payment is done by credit card, be associated with the customer's identification. These records could give the statistician fast access to precise data about the transactions and provide the economist with the statistics she needs for dynamic consumption analysis.

Toll roads have become common also in the Nordic countries. Application of the RFID technology is used to identify cars passing toll stations if the owners subscribe to an automatic payment system. From a small device on the front screen, the RFID reader at a toll gate forward, the passing car's identification to a central system in which it is time-stamped and the toll fare is subtracted from the car owner's account. In the case of a private owner, the account will be associated with her personal identification number as well as the car's licence number. Company cars are linked to the organizational identification. The records are saved for a certain time in case of complaints. If transferred to the NSI, the traffic data can be linked to data about the owner. These integrated statistical records can be important increments to statistics on use of cars, traffic patterns, etc.

Electronic travel tickets are frequently purchased over the Internet charging a bank account. In many cases, the bank accounts are associated with the official identification system. Access by the NSI to the electronic ticket transactions can provide a basis for statistics on air, land and sea traffic, travel patterns and migration, improving the present situation considerably, particularly if access to international immigration records is also granted.

4.2. Technological Trends

Technological trends point towards the development of small wireless transmitting and receiving electronic devices which can be attached to or embedded in artifacts. These devices can obtain the required operating power from their environment, be programmed remotely and receive instructions and transmit data interactively to communicating devices. The communication technology will vary from short-range radio

technology as the RFID, use of the mobile phone GSM nets, to GPS and low-orbit satellite communication (Wikipedia 2009a; IRIDIUM 2009).

A few examples indicate types of applications. Owners of cars and pleasure boats are now offered small boxes based on GPS, GSM and satellite technology that can be embedded in the car/boat and make it possible to track the whereabouts of the object. GPS is already an almost standard device for navigation in cars and boats. It is likely that in a few years the tracking box will be offered as a standard feature in new cars and boats with the option to subscribe to a tracking service storing tracking data. The system will at any time be able to provide the subscriber with electronic tracks for the object and its recent movements. Insurance companies may in the future offer reduced premiums to vehicle owners subscribing to such a service. Special insurance rates depending on where and when the vehicle is used may also be offered if the owner grants the insurance company access to the tracks of his vehicle. The tracking data will most probably be connected to official identifications of the vehicle and owner, and can become another new data resource for statistics on the use of cars and boats as well as travel statistics.

Corresponding devices exist for humans. In some countries, criminals can serve their prison sentences at home with restricted mobility. The condition is that the person is willing to wear an electronic *chain* around his leg. The chain contains a GPS unit tracking the person's movements, supplemented by a wireless transmitter reporting to a centre that he abides to the restrictions.

In hospitals, patients in intensive care can be monitored even when they are allowed to move out of bed by means of devices they wear and which communicate with a hospital control centre by means of a wireless local net. Out-patients needing frequent check-ups can be equipped with and carry similar types of devices communicating by means of GSM, relieving the patient from staying in a hospital and saving the hospital from spending unnecessary resources. Tracking devices are also offered to help locate old people having problems with finding their way back home.

We have become accustomed to the idea that we always carry our fingerprints with us as identification. Biometric passports for entry into a number of countries have recently been introduced (EU Council Secretariat 2007). To serve their purpose, the read identifications for people passing a frontier will be stored for some time. Advanced coding/decoding algorithms are used to reduce the risk of identity thefts.

Applications are not limited to devices installed or worn. Unique serial numbers have been *embedded* as part of cars for many years. *Implantable Medical Devices, IMDs*, are, as the name indicates, *implanted* in human patients in connection with serious illnesses. The pace-makers for heart function control are a well-known example. The IMDs, which are programmable, allow hospitals remotely to extend care to their out-patients (Fu 2009). *VeriChip* have already got their first chip for implantation approved in the U.S.A. (Wikipedia 2009b).

Will implantation in humans of chips emitting personal identifications be widely offered in the future? With an identity chip implanted, your driver's licence, passport, bank account, etc. can in principle be checked when you pass a reader with the required connection to the particular administrative system. In addition, the geographical position of the reader with a time-stamp can be stored with the rest of the data.

Implanting a chip in the human body could eliminate the need for carrying identity papers. As indicated by Figure 5, it would for example be possible for personal identities and those of the commodities carried in the shopping baskets to read automatically and the customers' accounts charged when passing an exit reader in the supermarket. With implanted chips, immigration officers at the frontier could be replaced by a portal screening your identity, your driver's licence could be checked even if you had left it at home, and the patient's medical history be automatically retrieved when the patient is brought through the entrance of the hospital emergency room.

These applications have a few things in common: they generate huge masses of data about individuals over time, they will probably be based on the official identification system and most of the data read will be considered sensitive. We may not like it and oppose the applications, but the technology exists and its use will most likely continue to increase. Because of its efficiency, experience indicates that the technology will be used and misused. But access to and analysis of such data can also revolutionize the conditions for certain aspects of social, economic and behavioural analysis and research, and provide trade and industry with data for more efficient operations. Obviously, access to the data will be considered very desirable, and the idea of National Data Centres is likely to surface again, followed by new intensive privacy debates (Albrecht and McIntyre 2005).

We shall refer to data discussed above as *Electronically Observed Data, EOD*, in contrast to the traditional data that are usually recorded with the active participation of the objects.

4.3. Implications for Official Statistics[4]

What implication may the increasing mass of micro data collected have for official statistics? What are the limits of the official statistics? Should an NSI refrain from using this growing stream of EOD in producing official statistics, or should it take advantage of the data to be able to provide the public with more information about the society? What would be an acceptable solution for balancing between the benefits of using these data to improve our knowledge of the society and the serious risk of misuse of the data? Is it likely that commercial interests will also see the value of and be interested in exploiting the data for commercial use? Should the NSIs engage actively in the development of a system that may reduce the risk of misuse to an acceptable level, and prepare themselves for the responsibility of being caretakers of EOD? Is it possible to design a legal framework regulating the use of sensitive data in a satisfactory manner, and can such regulations be effectively enforced? These are only a few of many questions official statisticians will have to consider in preparing their strategies for the future.

Two main future alternative scenarios can then be envisaged, one in which the NSI is not interested in nor entitled to access and use the EOD of administrative organizations, and a second scenario in which the NSI is permitted by a Statistical Law to access and

[4] The discussion in this section is limited to data stored with direct or indirect reference to the official identification system. In addition to data satisfying this requirement, there will be other data collected and stored for special purposes which may also be of interest to NSIs.

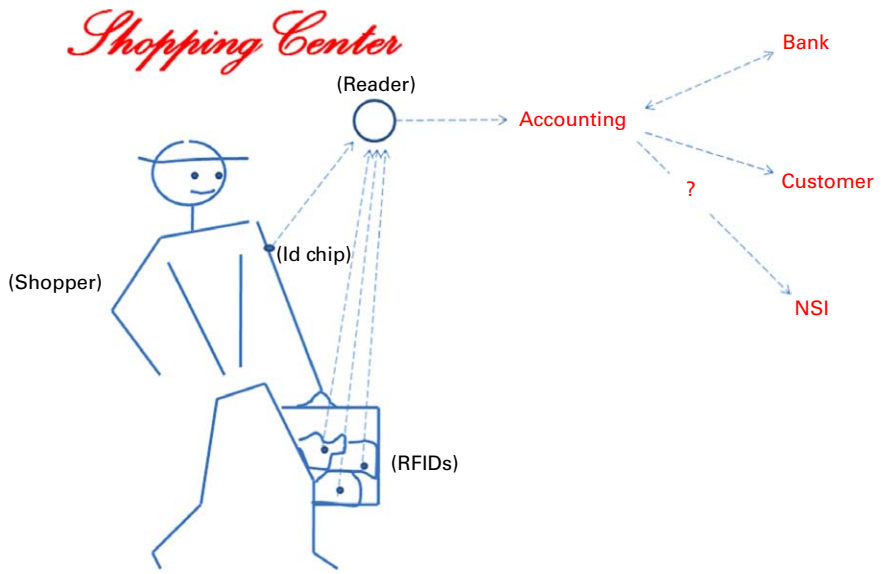

Fig. 5. A future shopper passing the cash reader

utilize administrative EOD in a manner similar to how it at present accesses and utilizes administrative data. The two alternatives are outlined in Figures 6 and 7.

Figure 6 indicates that according to the first scenario, administrative organizations, ADM 1 and ADM 2, will record EOD about objects for their administrative purposes and

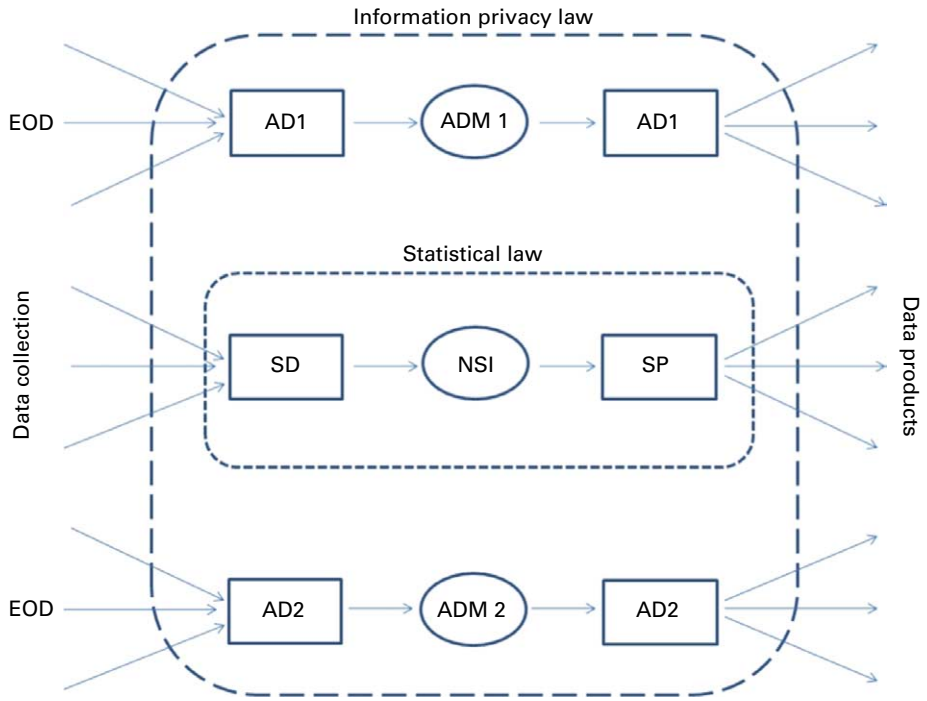

Fig. 6. Organization without NSI engagement in the new administrative data collection

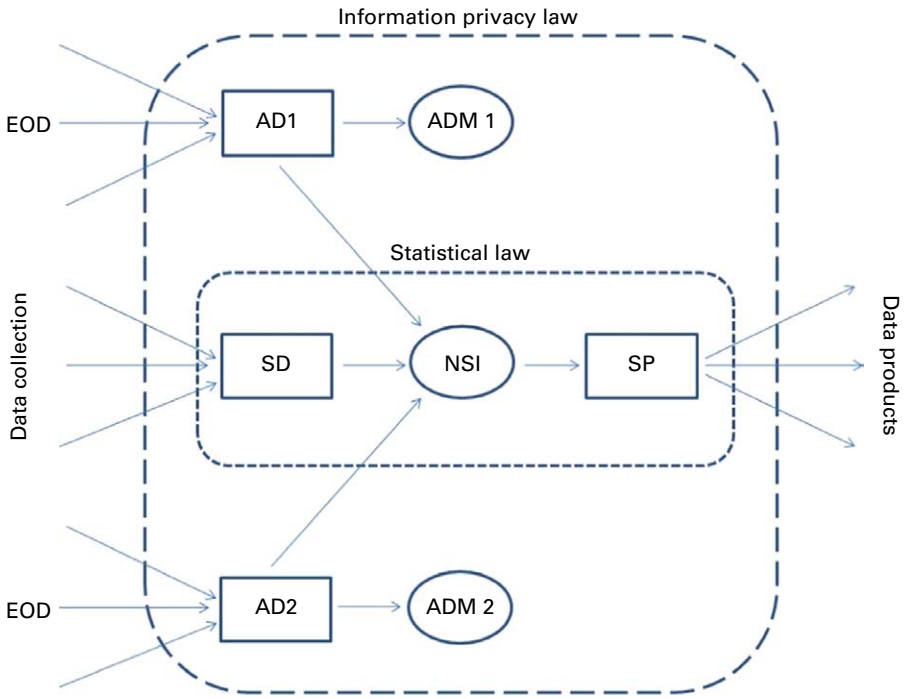

Fig. 7. Organization with NSI engaged in application of the new administrative data collection

probably, as far as the Privacy law permits, also make the data available for third party users. If the NSIs decide not engage in the use of EOD, official statisticians cannot be blamed for using administrative data considered too private or sensitive to be used for statistics. To collect statistical data comparable to EOD from administrative sources would be an expensive alternative for the NSIs if official statistics for the corresponding events would be required.

If the NSI decides that the use of the EOD collected for administrative sources is a desirable extension of the current use of administrative data for the preparation of statistical products, and is authorized by a Statistical Law to access and use administrative EOD as indicated in Figure 7, the NSI will, because of the official identification system, be able to integrate the EOD from different sources with its own data and provide better and more comprehensive statistical products without the costs of repeated observation of the objects.

In support of the engagement of NSIs as caretakers of EOD warehouses, the main arguments are that the EOD are too valuable to be wasted, the NSIs have well-established reputation for safe handling of confidential data, and the NSIs have the statistical and technical competence to obtain more useful information from the data than any other organization. The second scenario would require that the present statistical and privacy legislation be extended and adjusted to regulate the future situation for safe handling of EOD. The statistical legislation should give an exclusive right to the NSI for accessing, organizing and processing any data collected by other operators using the official identification system directly or indirectly. The NSI should be charged with responsibilities

for guarding the data against any access which can lead to misuse as specified by the legislation. The requirements would generate new methodological and technical challenges for the NSIs. In addition to the improved statistical methods required, technical methods for safeguarding the data against unauthorized use are the most important.

Since the National Data Centre proposal in the 1960s, requests from researchers for access to micro data sets for research have been increasing. Also an increasing number of business users want access to micro data sets for applying their own analytical methods to solving their particular problems, or presenting statistics for new users. Several approaches to by-pass the confidentiality problem have been evaluated and tested. Among the approaches tried are letting NSI staff members carry out the necessary processing according to the researchers' prescriptions, hiring the researchers as unpaid NSI staff members and as such subjected to the statistical laws, permitting the researchers to work with the micro data within closed and regulated processing systems, generating synthetic micro data set versions with the same distributional characteristics as the real data set, etc.

A challenge for the keepers of the EOD warehouse would be to develop practical and safe ways to prevent data and statistics for which they are responsible from being available and used in conflict with the privacy and statistical legislation. The general public must be informed and convinced that the developed solutions are effective and secure. It is important to be prepared for the future and to have a strategy for how to be prepared for the development. Some possible future technological trends of significance, some problems and possible solutions, have been indicated. It remains to analyse these in detail, draw the necessary conclusions and wait for the necessary political decisions.

To be prepared, the NSIs should put scenarios for possible future use of administrative EOD in the production of official statistics on their agendas for discussion, including requirements with regard to an adequate legislative framework and development of methods to ensure safe and proper use of EOD.

5. References[5]

Albrecht, K. and McIntyre, L. (2005). How Major Corporations and Government Plan to Track Your Every Move with RFID. Nelson Current. TN.

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

Biemer, P.P. and Lyberg, L. (2003). Introduction to Survey Quality. Business & Economics. Hoboken, N.J.: Wiley.

EU Council Secretariat (2007). www.eurunion.org/partner/euusterror/CounterTerrorFactsheetCoun030907.doc

Fu, K. (2009). Reducing Risk of Implantable Medical Devices. Communication of the ACM. Vol. 52. No. 6. pp. 25–27. doi:10.1145/1516046.1516055.

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

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

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.

Gåsemyr, S., Børke, S., and Andersen, M.Q. (2007). 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. (2008). Role of Editing and Imputation of Sources for Structural Business Statistics. Proceedings from UN/ECE Workshop Session on Statistical Editing in Vienna. Geneva.

Houbiers, M. (2004). Towards a Social Statistical Database and Unified Estimates at Statistics Netherlands. Journal of Official Statistics, 20, 55–75.

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. New York: Macmillan.

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, No. 2. Maiden, MA: Wiley Interscience. 223–234.

Lundell, L.-G. (2009). Strukturerade datalager för effektivare produktion. Unpublished paper. Statistiska Centrlbyrån, U/ARK. Stockholm. [In Swedish]

Miles, S.B., Sarma, S.E., and Williams, J.R. (eds) (2008). RFID Technology and Applications. Cambridge: Cambridge University Press.

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/

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. (1963). Automatic Editing of Individual Statistical Observations, Statistical Standards and Studies, No. 3. United Nations, N.Y.

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

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.

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 Canada (2009). History of the Census. Ottawa. Available online at http://www.statcan.gc.ca/edu/power-pouvoir/ch2/history-histoire/5214912-eng.htm

Statistics Denmark (1995). Statistics on Persons in Denmark – A Register-based Statistical System. Eurostat, Luxemburg.

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. Stockholm: Statistiska Centralbyrån.

Sundgren, B. (2004). Statistical Systems: Some Fundamentals. Statistiska Centralbyrån, Stockholm. Available online at http://sites.google.com/site/bosundgrenknowledgebase/

Sundgren, B. (2010). The System Approach to Official Statistics. In this book.

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

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

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

Wallgren, A. and Wallgren, B. (2007). Register-based Statistics: Administrative Data for Statistical Purposes. Hoboken, N.J.: Wiley Series in Survey Methodology.

Wikipedia, English Edition (2009a). RFID Technology. http://en.wikipedia.org/wiki/Radio_frequency_identification

Wikipedia, English Edition (2009b). VeriChip. http://en.wikipedia.org/w/index.php?title=Verichip

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.