# USE OF ELECTRONICALLY OBSERVED DATA IN OFFICIAL STATISTICS[1]

Nordbotten, Svein
*Svein Nordbotten & Associates*
*Rieber-Mohns veg1*
*5231 Paradis, Norway*
*E-mail: svein@nordbotten.com*

## 1. Official statistics

An industrialized country in the 21st century can be considered as a complex system of interrelated objects acting more or less according social 'laws'. To make efficient decisions within a modern society will require statistical information about these laws the preparation and dissemination of which is the main objective of the National Statistical Institutes (NSIs).

In this presentation, two sources of data collected by the NSIs for preparing official statistics are considered:

1. *Administrative data* collected by other organizations in performing their special duties, and submitted to the NSI for statistical utilization, frequently regulated by statistical legislation.
2. *Statistical data* collected by the NSIs solely for statistical use if no adequate administrative data are available.

The use of administrative data has become increasingly important during the last decades because of 3 reasons:

1. effective electronic transfer of data from administrative to statistical agencies,
2. respondents' demand for reduced response burden and,
3. restricted budget grants to NSIs.

Official statistics have continuously improved in details and geographical coverage as well as in quality. In those countries which have a public administration based on a system of shared population, business and property registers with permanent and unique identifiers, the NSI can now substitute statistical data previously collected by statistical surveys and censuses, for e.g. population censuses, with data collected from different administrative sources and integrated to the required survey/census files [UN/ECE 2007a]. By integration and reuse of administrative data, new statistics

[1] Prepared for presentation at the 58th World Statistics Congress of the International Statistical Institute, Special Topic Session 25: Analyzing Internet Traffic Flows and digital footprints for statistical purposes, Dublin August 2011.

are also developed in areas previously uncovered by official statistics [Nordbotten 2010a, Nordbotten 2010b].

## 2. Technological trends

The improvements in official statistics achieved during the last 50 years would have been impossible without the electronic information technology which supplied the statisticians with effective tools to process and organize the accessible data.

Today, new technologies such as:

1. magnetic stripe/microchip identification (Credit cards, passports),
2. radio frequency identification (RFID),
3. mobile phone networks (GSM),
4. global positioning system (GPS),
5. communication satellites (COMSAT),

can create new data sources for official statistics. These technologies permit objects to be observed electronically and recording events at the time they occur without significant response errors or burdens for the respondents.

Applications of the new technologies include for example recording of economic transactions, objects passing immigration controls and toll-road gates, and tracking of the location of vehicles and individuals over time. The use of credit/debit cards with *magnetic stripes/microchips* completing economic transactions is well known in most countries. A number of countries have introduced biometric passports permitting electronic recording of identity and demographic attributes of travelers passing country borders/passport zones.

*RFID tags* are used in cars for automatic and instantaneous identification of vehicles passing RFID readers located in automatic toll-road stations, garage gates, etc. Similar tags can be embedded in commodities substituting bar-codes for recording identities of commodities in shops, warehouses and factories. RFID tags are also carried by humans to identify their access to restricted areas, and have even been used by conference organizers to study attendance patterns.

*GPS devises* combined with *GMS* or *radio satellite communication*s are used for tracking the positions of boats and cars carrying such devices. Insurance companies may in the future offer reduced fees for vehicles equipped with this technology to help discovering lost or stolen objects. The technology also contributes to increased personal safety both on sea and land.

Similar devises can be carried by humans and animals. For example, criminals are allow to serve their sentences at home if carrying tracking devises, and patients needing continuous or frequent

monitoring can carry similar devices and be sent home saving hospital costs. Miniaturized versions can even be permanently implanted into their bodies.

The cost of electronic observation recording is decreasing, new applications are emerging continuously, and the amount of Electronically Observed Data (EOD) generated per time unit is growing.

## 3. New possibilities for official statistics

Electronic observations are cost-effective. Most of the current applications were established for administrative purposes. The collecting organizations use the EOD for control and statistical analysis of their operations.

The EOD are frequently representing events of importance for the domain of official statistics, they are already in compact electronic form and inexpensive to reuse, observation errors are rare, statistical reuse does not require any burden for the 'respondents', and the EOD represent events which would otherwise be difficult to observe. Obviously, there is a potential for reuse of the EOD in preparation of official statistics. So far, few proposals for reuse of EOD in official statistics have been promoted and discussed, mainly because much of the EOD are sensitive and present legislation restricts the transfer, storage and reuse of the data.

Future EOD reuse applications may follow along one of alternative paths:

1. One extreme scenario is that saving and reuse of EOD will not be permitted because of privacy, business interests and risk of misuse.
2. The other extreme alternative is that the demand for reuse will force the market value of EOD to increase and primary collectors will be permitted to freely dispose of their EOD.
3. A third scenario is that respect for privacy, business interests and the need for preventing misuse will be acknowledged at the same time as a solution for safe reuse will be established.

We assume the third alternative will prevail, and that the NSI, as for other administrative data, will be charged with the responsibility to access, organize, save and reuse EOD subject to strict and safe confidentiality rules. The reuse will comprise the preparation of regular statistics, as well as *ad hoc* reuse by means of EOD for social and economic analysis and research. In countries were the NSI is an organization already trusted with storage and reuse of administrative microdata for preparation of statistics, NSI will be an ideal organization also for safe storage and reuse of EOD.

The EOD cover already many statistical domains of interest and being independent of manual collection and preprocessing, EOD can be a basis for continuous and highly timely statistics. In countries with public identification systems, the EOD are frequently linked to the official

identification number system. Integration with other object data and reuse for preparing new, more detailed and timely statistics of high value for as well government as private business and the general public can be envisaged. Of particular interest are the possibilities to create time series of microdata which can form a basis for valuable longitudinal studies of the dynamics of the society.

A couple of examples can outline the possibilities. Consider a chain of food shops which has substituted the bar codes with RFID tags (See Figure 1). A customer comes to the cashier where a reader automatically identifies the selected goods in his cart and a list of the goods, their prices and the total for payment (lines 1 and 3) are displayed. The customer pays with his credit card, a copy of the transaction is sent to the NSI (line 3) and his bank account is charged (line 4).

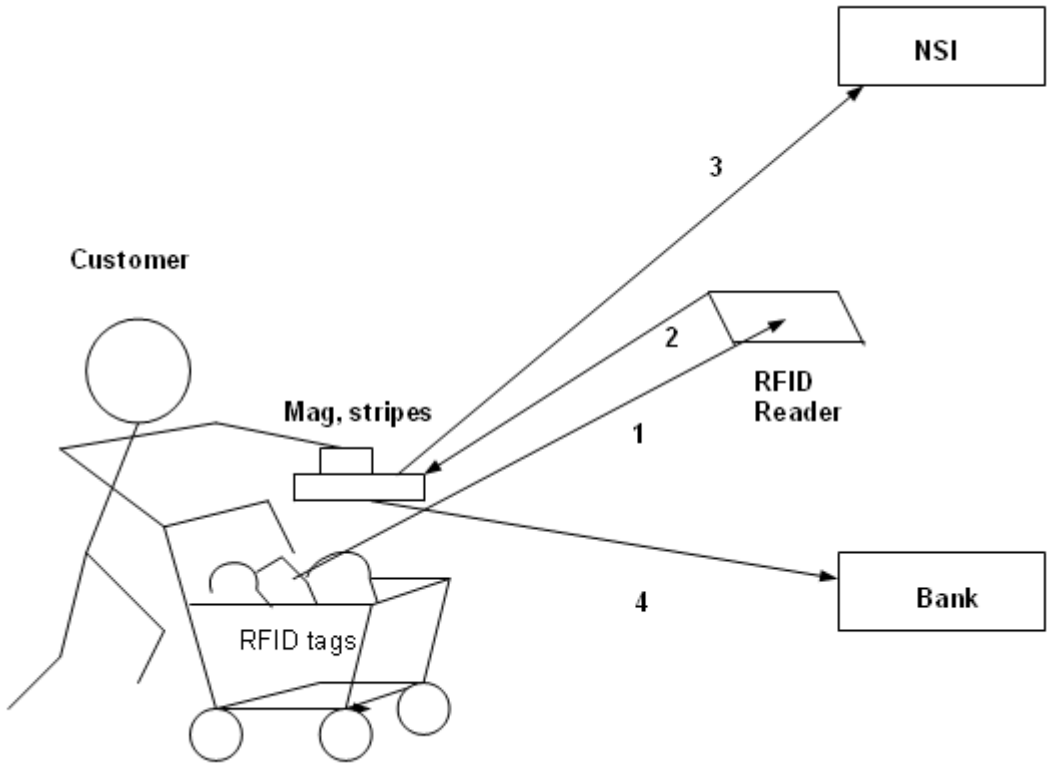


*Figure 1: Shopping transactions*

Timely sales statistics can be prepared for different commodities and commodity groups by geographical areas and by time of day. Because the EOD in principle can be available immediately after the transactions, up-to-data statistics of market trend indicators can be made available for the government, businesses and general public on a daily or weekly basis.

In countries with public identification systems, the transaction EOD can in principle be integrated with other attributes for the customers and the shops permitting the sales also to be broken down by the income group, profession, age, etc. of the customers and by size, category and location of the shops. By generating individual time series records, longitudinal research contributing to better

understanding of the factors determining market development including the decisions of customers and shop managers can be obtained.

Assume that insurance companies are offering rebates if owners install tracking devices in their vehicles and that the EOD from tracking are made available for the NSI (See Figure2). The device installed in the vehicle receives signals (line 1) from GPS satellites from which the vehicle position is computed. The vehicle identification and position are frequently sent via communication satellites to the insurance company (lines 2, 3 and 4), and copies can be accessed by the NSI (line 5). These EOD can be used for preparing detailed statistical tables of traffic and transport pattern statistics by time and local areas with high flexibility in specification of time and areas.

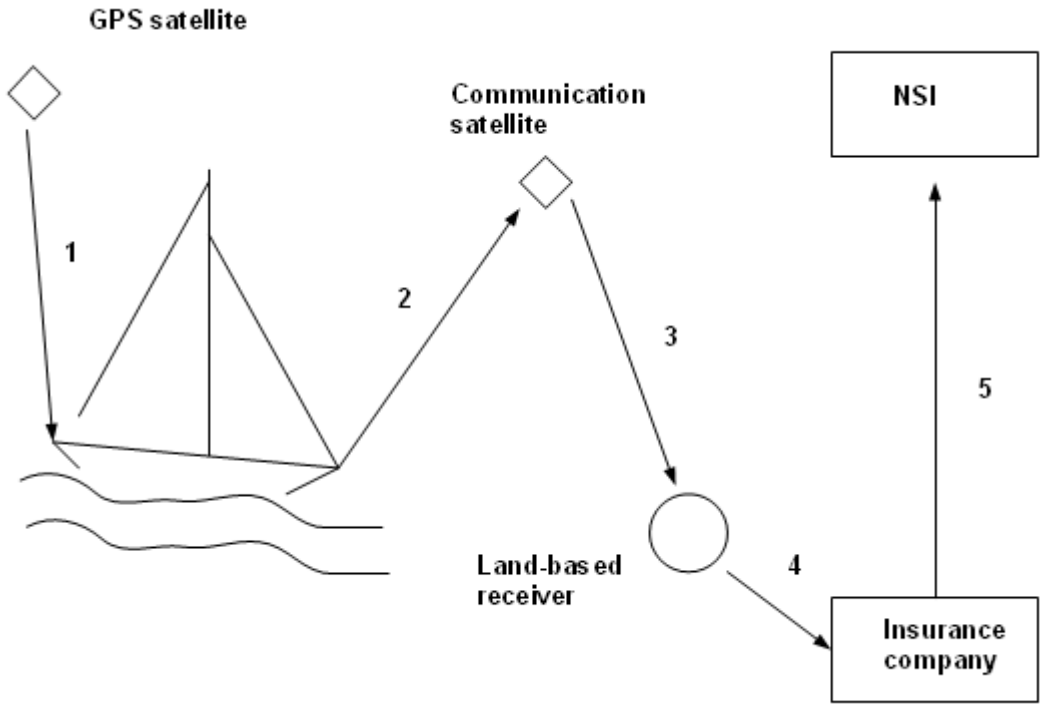


*Figure 2: Vehicle tracking*

In countries in which the vehicle registration is part of a general public identification system, tracking data can be integrated with recorded data for the vehicles and their owners permitting the preparation of statistics for an even wider range of demands. As in the previous example, the EOD can be integrated into microdata time series which can be used for longitudinal studies of travelling patterns, etc.

So far, we have been focusing on reuse of ' administrative' EOD. Inexpensive devices with GPS and communication facilities can also be utilized directly by NSIs in their own statistical collection. Collection of EOD for statistical purposes can be implemented by making agreements with a sample of people to carry the devices for a certain time reporting automatically for instance their

commuting patterns. These devices can be considered as robot interviewers/observers who can be used in statistical surveys and, if wanted, combined with traditional collecting methods.

The EOD can be supplementing, in some cases also substituting, the present administrative and statistical data collection for official statistics. Because of the NSIs' good reputation, statistical competence and experience, no other organizations can be trusted as effective and reliable re-users of EOD for statistical purposes.

The precise, fast and automatic recording of EOD make the data representing information of high value in addition to the value for the operations they were originally initiated. Since the recording is carried out in real time, re-observation would be expensive, and less accurate. The balance between the value of EOD for statistical reuse on one side and the risk of misuse on the other requires therefore careful consideration.

Even when administrative data required for a certain statistical processing exist and are available for a NSI, it happens that a statistical collection is still executed because the administrative data are not considered adequate. The data may not reflect the concepts wanted, the population covered by the administrative operation is not as wanted by NSI, the data collected are considered incomplete, the collection procedure generates too many errors, and/or the data are not accessible fast enough. Similar objections can also be raised against reuse of EOD from some administrative applications for preparation of official statistics.

## 4. Confidentiality implications

Resistance against the collection and storage of personal data was significant 50 years ago [Miller 1967]. The first protests against collection of EOD appeared also several years ago based on privacy considerations [Miles 2008]. Objections against reuse of EOD in preparation of official statistics must be expected. In most countries, official statistics and the NSIs are, however, highly respected and objections will probably not be against the statistics generated, but against the risk of leakage and misuse of sensitive EOD. It will be a major task to convince the public that the statistical benefits justify the risks of misuse when they are kept within acceptable limits [Sundgren 2001].

Legislation may be needed permitting reuse of EOD for official, statistical purposes by the NSI and specifying the security conditions required. This must imply that the NSI is granted the privilege to access, store and process such data subject to strict confidentiality conditions including a tight security wall around the NSI.

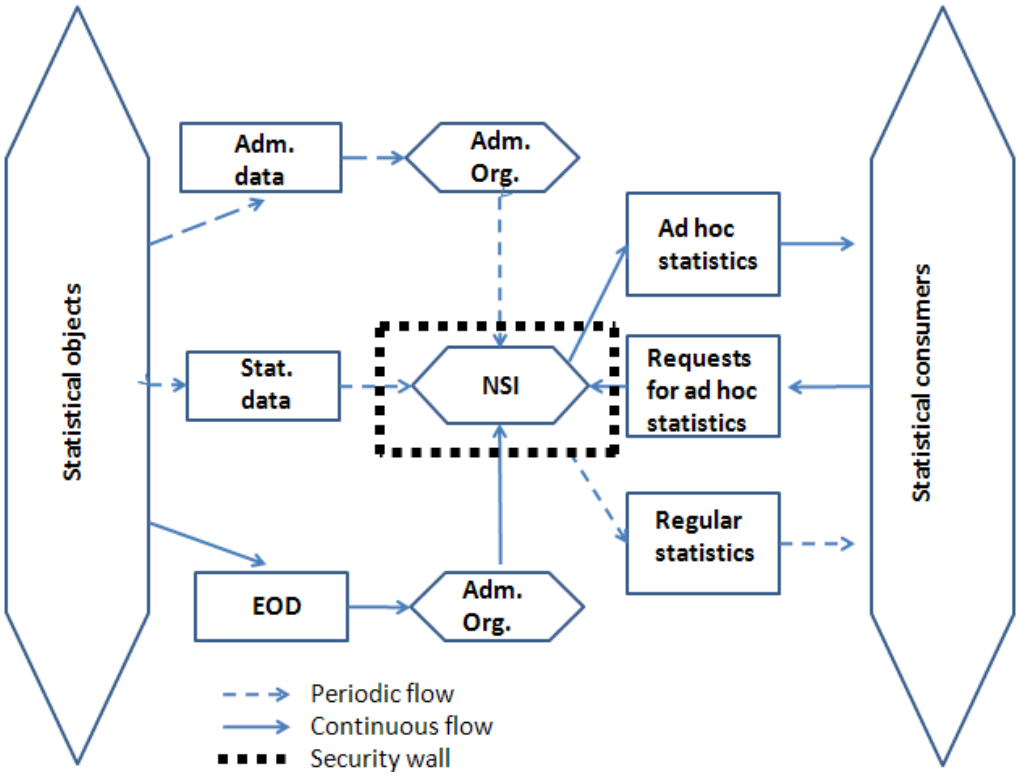


*Figure 3: Relations between objects observed and statistics provided*

Figure 3 illustrates how administrative data, statistical data and EOD can be compiled in a NSI and used in preparing regular statistics and *ad hoc* statistics on-demand. To maintain the public trust, a major point will be that the security wall will permit microdata to flow in, but not out of a NSI, i.e. all requested preparation of *ad hoc* statistics has to be executed within the NSI and the reported results checked for confidentiality before sent to the users [UN/ECE 2007b].

## 5. Technical challenges

Reuse of EOD by NSIs presents also new technical challenges. Each EOD record will at least contain identification of the recording device which implicitly specifies the type of event observed, identification of the observed object, the observed location and a time stamp. Depending on the administrative purpose, several basic records can be linked in a batch. To take advantage of the real time recording of EOD for fast statistical reporting, the NSIs must develop capability to access and store parallel streams of data from the collecting administrative organizations. A NSI can decide that a complete copy of an accessible steam is not needed, and develop methods for sampling event records from the stream.

When possible, the received EOD should also be integrated with other data for the observed and related statistical objects according to a predesigned time schedule. The integration must aim at both latitudinal and longitudinal microdata records to be prepared to serve a variety of demands for *ad hoc* statistics.

Statistical data must be stored in databases designed to respond fast and effectively to remote and on-site demands for datasets required by a diversity of applications. Many NSIs already have well developed databases serving queries for already existing macrodata. Well designed databases for microdata will be needed to respond to demands for statistical information which require reuse and new integration of microdata. Such *ad hoc* processing will require methods for serving remote as well as on-site requests for processing and for making certain that the outputs satisfy the confidentiality conditions.

## 6. Conclusions

Based on the above discussion the following conclusions are suggested:

1. Data collected by means of electronic devices are currently being generated in an increasing number of administrative applications. Reflecting important activities in modern societies, stored EOD represent valuable information, but also a risk for misuse.
2. The NSIs should discuss their interest in collecting and processing this type of data for preparing official statistics.
3. If a NSI decides that the EOD from administrative applications can be an important, future source of data for preparation of official statistics, the NSI should start preparing for future use of EOD.
4. Getting access to and utilizing this new data source will probably require new or extended legislation. The NSI should engage actively in preparing a legislation which grants the NSI full and immediate access to EOD and responsibility for organizing and safe reuse of the data.
5. This responsibility will require development of methods and systems including security walls, accessing and sampling data streams, integrating EOD with other microdata, using the integrated records of microdata to improve data quality, storing microdata in adequate databases, fast retrieval and processing response to on-demand request for *ad hoc* statistics, preserving confidentiality and providing remote processing service.

## REFERENCES (RÉFERENCES)[2]


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

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

Nordbotten, S. (2010a):*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. 205-225. Available at: officialstatistics.wordpress.com/

Nordbotten, S. (2010c): *The statistical archive system 1960-2010: A summary*. Nordisk Statistikermøde i København 11.-14. august 2010. København.

Sundgren, B. (2001): *Statistical Microdata –Confidentiality Protection vs Freedom of Information*. Joint ECE/Eurostat Work Session on Statistical Confidentiality. UN/ECE and Eurostat. Available at: http://www.sites.google.com/sites/bosundgren/

UN/ECE (2007a): *Register-based Statistics in the Nordic Countries – Review of Best Practices with Focus on Population and Social Statistics*. United Nations. N.Y/Geneva. Available at: http://www.unece.org/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

UN/ECE (2007b): *Managing Statistical Confidentiality & Microdata Access – Principles and Guidelines of Good Practice*, United Nations. N.Y./Geneva. Available at: http://www.unece.org/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf


[2] Articles by the author are available as full text copies at http://www.nordbotten.com/articles/.

**RÉSUMÉ (ABSTRACT)**

*National Statistical Institutes (NSIs) collect electronic data from administrative agencies as well as directly from individual respondents. As a next step, modern technologies such as radio frequency identification, mobile telephone nets, geographic positioning systems, low orbit satellites, etc. also permit electronic observations without active participation of statistical objects observed. A number of applications and services have already been launched such as wireless recording of cars passing road toll gates, tracking of whereabouts of boats, dogs and individuals. Electronically Observed Data (EOD) represents a new and valuable addition of data reflecting current activities in a complex society. The EOD are, however, frequently sensitive data. Preserving and utilizing EOD is likely to become central topic in the years to come followed by discussions about how to regulate access to satisfy an acceptable balance between benefits and risk of misuse. In particular, NSIs in countries like the Nordic with well developed central identification registers for population, enterprises, organizations, vehicles, properties, etc. will probably experience that the EOD uses are directly or indirectly connected to their own data by their central identification systems, and be facing a demand for special services including EOD. The NSIs have long traditions in confidential treatment of sensitive micro data, and should consider and prepare for their roles and responsibilities with respect to EOD.*

.