



Use of non-survey data in official statistics

Roger Jensen and Arild Langseth, Statistics Norway

1. Abstract

In Statistics Norway's strategy, one of the main goals is to collect, use and share data for the benefit of the entire society. Furthermore, the written guidelines from the Ministry of Finance emphasize that Statistics Norway shall utilize the growth in new data sources for statistical purposes, and that new data sources can contribute to realizing Statistics Norway's focus areas.

In this paper we focus on Statistics Norway's use of existing data sources that were not originally intended for official statistics, but which can be processed and reused for such purposes. This includes both administrative and other data and is referred to in international literature as "non-survey data". Such data differ from data collected via statistical surveys.

First we give an overview of non-survey data that can be useful for official statistics, along with key aspects most relevant to this discussion. Then we summarize Statistics Norway's most important experiences so far with the collection of non-survey data. Finally, we outline the main directions for the further work with the use of such data.

2. A brief overview of data sources

Non-survey data exist regardless of whether they are used for official statistics or not. Instead of asking respondents what they do and think, these data sources open up the possibility of producing the statistics based on the traces left by their actions and attitudes; whether it is process data or data about products and services.

TABLE 1 Non-survey big data sources by type

Type of source	Example of data
Register	Vital events, diagnoses Wage, income tax, VAT, welfare payments
Transaction	Scanner data price, point-of-sales receipt Bankcard or giro payment B2B or B2P invoice Property sales contracts, ownership registration
Remote sensing, fixed	Smart meter readings Weather station readings Traffic loop signals
Remote sensing, mobile	Satellite images, drone images Airborne laser scanning Maritime AIS, lorry tracking signals Mobile phone signals
Internet	Web pages Social media posts

The table above (Zhang and Haraldsen, 2022¹) provides an overview of different types of non-survey data. While Statistics Norway has collected administrative data from government agencies and other public data owners for a long time, commercial companies are often owners of other non-survey data sources. However, the benefits of these data from the business sector are the same as of administrative data, i.e. to reduce both existing and future burden on businesses, develop new statistical products and more detailed statistics, and improve timeliness where possible.

Replacing traditional data collection with non-survey data could reduce the total burden on individuals and businesses. However, such changes will shift the burden from individuals or businesses to third-party actors. Good cooperation is crucial for these actors to prioritise the necessary resources for the establishment of data collection solutions. Furthermore, cooperation and good routines are also needed to make sure that the data collection can run smoothly over time.

Two common features of such data and their interaction are particularly important:

- 1) **Quality:** Since the data were not originally intended for official statistical purposes, data from some sources often need to be combined and processed together with other available data, e.g. to establish relevant statistical units, define populations, derive or estimate variables, etc. Without a combination of sources, a single data source can rarely satisfy the quality requirements of official statistics.
- 2) **Consent:** The traditional, informed consent in Statistics Norway's own surveys does not exist with the owners of such data. Instead, Statistics Norway must address privacy and other cost considerations for data subjects through cost-benefit assessments and data protection impact assessments. Privacy and data protection measures must be applied both when collecting and processing data during the production of statistics, and Statistics Norway must constantly create understanding through information work aimed at the general public.

3. Status

The status of collection and use of different types of data in Statistics Norway is as follows:

Register data

Statistics Norway has a long tradition of using registers and other administrative data for official statistics. In most cases, this takes place through formalised agreements on the delivery of data to Statistics Norway and the cooperation on quality in administrative information systems. Since 2012, Statistics Norway has continuously entered into such cooperation and now has 29 agreements. In the agreements, around 190 different data deliveries are specified. Data deliveries vary widely, both in frequency and scope. Some are received daily, while others can be received weekly, monthly, quarterly or annually. Some consist of millions of units and hundreds of pieces of information about each unit, others of just a few units with limited pieces of information.

In addition, Statistics Norway receives register data from a number of other enterprises, both public and some private, without any special agreements as mentioned in the section above.

Modernisation of registers and information systems is currently taking place in many public agencies. This means that Statistics Norway will increasingly be able to use new data collection solutions in the future. In some cases, this may mean that more data collection is done more frequently and perhaps

¹ Zhang, L.-C- and Haraldsen, G. (2022). Secure big data collection and processing: Framework, means and opportunities. *Journal of the Royal Statistical Society, Series A*, <https://rss.onlinelibrary.wiley.com/doi/10.1111/rssa.12836>

also that more detailed data are made available to Statistics Norway, which can result in better and more relevant statistics.

Transaction data

Statistics Norway has experience with scanner price data from various industries, sales receipt data from grocery chains and card transactions from acquirers. Sales receipt data and card transactions at transaction level have great potential for use in statistics on individuals, but the processing of the data also raise privacy issues. The collection of such detailed transaction data has therefore been temporary suspended. On the other hand, aggregated card transactions still have considerable statistical value in business statistics, considering that privacy challenges are minimal or non-existent. Statistics Norway therefore collects both debit and credit card transactions aggregated by merchants on a daily basis. In addition, Statistics Norway is also considering giro transactions as a data source for business statistics.

Remote sensing, fixed

One example is electricity consumption and electricity production based on smart meters. Statistics Norway receives detailed data for all smart meters in Norway aggregated by monthly frequency. These are data that are used in electricity statistics and may also have potential for use in other areas, e.g. to determine if a building is being used or not. Statistics Norway also aims to develop statistics based on vehicles passing electronic toll gates.

Remote sensing, mobile

One example is the ongoing work to use remote sensing to monitor construction that may lead to the destruction of natural areas. Here we look at whether satellite images can help us determine when the construction took place. Furthermore, Statistics Norway has had an initial dialogue with several of the mobile network operators in connection with the potential use of mobile positioning data. Aggregating data can be a solution to avoid privacy challenges. Statistics Norway is now participating in a project together with several other countries financed by Eurostat, where the purpose is to look at how aggregated mobile data can be linked to other data sources for use in official statistics.

Internet

Typical examples so far are webscraping of price data.

4. Experiences/lessons learned

The following are the most important experiences Statistics Norway has made so far:

- Statistics Norway has extensive experience with the use of administrative data and registers in the production of statistics. Although this data was not originally intended for official statistics, it has proved to be well suited for the purpose. The formalised agreements on the delivery of data and on data quality with other public agencies has contributed to a better control over the data deliveries and the quality of the data received. The vast majority of deliveries of administrative data and registers are delivered as agreed, and it is very rare that delays or problems with the data lead to problems in the production of statistics.
- The principle of equality is important for commercial data owners, i.e. the duty to deliver data should preferably be imposed on all data owners rather than on a few.
- There is a significant privacy threshold for collecting data on individuals from private companies, such as in the case of sales receipt data from the grocery chains and mobile positioning data from mobile network operators. The grocery chains Coop and NorgesGruppen appealed against the decision on the duty to deliver data and asked The Norwegian Data Protection Authority for

advice. The Norwegian Data Protection Authority considered the case and finally decided to prohibit Statistics Norway from processing sales receipt data from the grocery chains to the extent as planned. Statistics Norway decided not to appeal this decision, but instead have a dialogue with The Norwegian Data Protection Authority about how Statistics Norway can utilize this type of data source. As for now, the collection of such detailed transaction data has been temporary suspended. The mobile operators Telia and Telenor have so far been sceptical about providing mobile positioning data to Statistics Norway at other than aggregated level. Statistics Norway is therefore now focusing, in cooperation with other countries, on aggregated mobile positioning data and how these data can be linked to other data sources for use in official statistical production.

- Data quality is not always sufficient for statistical purposes, and it may require an effort to raise the data quality, especially information that is less important to the data owners, whether it is administrative registers or other types of data.
- Linking with Statistics Norway's basic registers or other data sources is usually necessary to achieve sufficient quality for use in official statistics. One example is linking sales receipt data to card payments to be able to append household information. Another example is linking merchants (which receive card payments) to Statistics Norway's Business Register in order to define the population and associated industry groups, as well as to add other relevant variables.

In particular, there is a notable trade-off between needs-driven versus opportunity-driven development when Statistics Norway uses data from commercial data owners:

- Needs and benefits are most evident in business statistics and economic statistics. Together, different types of transaction data (Appendix A) and other non-survey data are likely to replace some of today's sample surveys, except from the most important units, and data can be made available earlier and be more detailed at the same time. For example, the current web survey for the Index of wholesale and retail sales may in the long term be partly replaced by various types of transaction data.
- For social- and household statistics, the motivation for using non-survey data usually lies in the possibility of new statistics. For example, information in sales receipt data from the grocery chains can be used in new statistics on diet. Mobile positioning data can be included in the data basis for new statistics on work commuting, tourism, travel and leisure activities (e.g. Analysys Mason, 2021²), but mobile positioning data cannot directly replace current statistics in these areas and the associated sample surveys.

5. The way forward

Further work on the use of non-survey data in Statistics Norway is being pursued along two lines:

- In order to rapidly realise the benefits of new data sources, it seems most effective to focus on business statistics and economic statistics, i.e. to utilize as many sources as possible for this purpose as quickly as possible. Examples of such sources are:
 - Card transactions aggregated by merchant.
 - Giro transactions at enterprise-level.

² Analysys Mason (2021). Nordic Development Project on Mobile Positioning Data for Tourism Statistics. <https://www.nordicinnovation.org/programs/mobile-positioning-data-tourism-statistics>

- Methods to minimise privacy and data protection concerns for individuals and at the same time make data better suited for statistical production, e.g. by combining sources. This work is most relevant for social- and household statistics.
 - When combining methodology and technology we can design new methods where data can be linked and used for statistics while privacy is protected (Zhang and Haraldsen, 2022).
 - At the same time, this presupposes that we can create an understanding and acceptance of the data collection through information work aimed at the general public.
 - The development of new systems, technology and methods will lead to additional costs. Cost-benefit assessments must be carried out to decide which solutions and methods to develop.

References

Analysys Mason (2021). Nordic Development Project on Mobile Positioning Data for Tourism Statistics.
Zhang, L.-C- and Haraldsen, G. (2022). Secure big data collection and processing: Framework, means and opportunities. *Journal of the Royal Statistical Society, Series A*

Appendix A. Transaction data for the use in business statistics and economic statistics

Both payment instruments and, not least, data sources are constantly evolving and changing. The overview must be maintained continuously and the system for data collection adjusted accordingly, if transaction data are to replace a large part of today's sample surveys in economic statistics.

