



Economic Commission for Europe

Conference of European Statisticians

Seventieth plenary session

Geneva, 20-22 June 2022

Item 9 of the provisional agenda

Collaboration with private data providers**Collaboration with private data providers at national level****Experiences and challenges with access to privately owned data about mobile phone positions in Sweden****Prepared by Sweden***Summary*

Mobile network operators (MNOs) capture, process and store data about the geographical positions of mobile phones every day. Anonymized and aggregated data derived from these phone positions can be used for commercial exploitation of mobility patterns and producing official statistics about population mobility, commuting, tourism and long-distance travelling. With this in mind, Statistics Sweden and the biggest telecom operator in Sweden (Telia) started to collaborate in 2020 to explore how statistical offices and MNOs which exploit these data commercially can reinforce each other. The paper discusses legal and practical obstacles and required steps to get guaranteed long-term access to MNO-data.

The document is presented to the Conference of European Statisticians' session on "Collaboration with private data providers" for discussion.



I. Summary

1. People generate billions of data points about the geographical positions of their mobile phones every day. Mobile network operators (MNOs) capture, process and store these data for business purposes such as increasing the network performance. Anonymized and aggregated data derived from these phone positions (MNO data) can be used for commercial exploitation of mobility patterns and producing official statistics about population mobility, commuting, tourism and long-distance travelling. With this in mind, Statistics Sweden and the biggest telecom operator in Sweden (Telia) started to collaborate in 2020, one month before the outbreak of the COVID-19 pandemic. The business model was 1) deliveries of aggregated and anonymized data to Statistics Sweden, 2) in exchange for knowledge exchange and benchmarking with Statistics Sweden's statistics.

2. Interest from the public and government agencies in MNO data increased after the outbreak of the pandemic because they clearly showed changes in population movements. At the same time, it appeared from a statistical perspective that creating stable time series from MNO data is not straightforward in periods of sudden change, hampering exact quantification and producing official statistics out of it. Well-described processes and methodological improvements were needed to tackle this issue. The collaboration with Telia gradually changed into a knowledge exchange partnership with Statistics Sweden position as passive and advisory in descriptions of data-processing and geolocation but active in the 'weighting' issue: relating mobile devices to human populations.

3. A government assignment followed for 2021-2022, which Statistics Sweden carries out with two MNOs and other governmental agencies. The aims of the assignment are 1) to describe conditions for use of MNO-data for official statistics, 2) to develop new 'smart' statistics, and 3) to replace (parts of) current surveys with these data. The government assignment changed the knowledge exchange partnership into a more general public-private partnership. In this partnership, it is explored how statistical offices and MNOs which exploit these data commercially can reinforce each other. The aim is to continue to collaborate with MNOs after the government assignment. The paper discusses legal and practical obstacles and required steps to get guaranteed long-term access to MNO-data.

II. Background

4. Four mobile phone network operators (MNOs) are active in Sweden. The biggest one (Telia), which is also active in Denmark, Finland, Norway and the Baltic States, commercially exploits its aggregated and anonymized mobile phone position data (MNO-data) by the so-called Telia Crowd Insights platform. Telia Crowd Insight does not cover Sweden alone. The platform commercially exploits MNO-data for all Nordic countries (excl. Iceland) and the Baltic States. Telia Crowd Insights drew general attention from governmental agencies and the media during the COVID-19 pandemic as these quickly available MNO-data showed plausible changes in population movements after the outbreak of the pandemic. The pandemic also demonstrated the potential of MNO-data for official statistics and policy analyses. However, the pandemic also showed quality issues such as transparency, documentation and a methodological one: relating devices to humans. These issues must be resolved before using these data for official statistics. As Telia also wanted to resolve these issues, the collaboration gradually changed into a partnership. After a government assignment about the use of MNO-data for official statistics in 2021, Statistics Sweden also started a partnership with a second operator (Tre), this MNO has a similar but less documented process than Telia.

5. These two MNOs together share approximately 50 % of the market. Telia's market share is close to 40 %, Tre has a market share of slightly over 10 %. These market shares are approximately constant over time since 2019. Regional differences in market shares do exist, partly because Telia only has a network which covers remote parts of Sweden completely.

III. Preconditions for collaboration and partnerships

6. Data privacy and GDPR (General Data Protection Regulation) are important issues when dealing MNO-data. All MNOs in Sweden apply high standards regarding data privacy. They interpret GDPR rigidly to prevent image damage at all costs. The consequence is that MNOs follow individual devices only for a limited time period (24 hours in the case of Telia). Another consequence is that client information is not used in the process. The third consequence is that MNOs are willing to deliver aggregated data to Statistics Sweden only together with process documentation.

7. Several legal issues had to be resolved before receiving aggregated data and processing documentation. These issues are related to 1) the Equal Treatment Principle, 2) the Open Government Principle and 3) statistical confidentiality.

8. A lesson learned is that it was time-consuming to contact all MNOs after the reconnaissance study in 2020 and provide them with the same collaboration offer. This procedure was necessary due to the Equal Treatment Principle, because MNOs may profit from Statistics Sweden's statistical knowledge when collaborating. The Open Governmental Principle implies that all shared documentation between Statistics Sweden and MNOs is public. Shared documentation between MNOs and Statistics Sweden may, however, reveal the confidential business processes of an individual MNO. On the other hand, the statistical law prevents and forbids revealing information about individuals and individual enterprises. In this case, MNOs were willing to share documentation about their processes, but requested that their process descriptions should be treated as confidential. Although this request was supported by SCBs statistical researchers, descriptions of MNO processes may indeed reveal confidential information about individual MNOs. A lesson learned is that this request appeared to be complex. It resulted in discussions with legal experts under which conditions the delivered information about MNO's data-collection processes are covered by the statistical law. Consequently, Statistics Sweden cannot share all information about MNOs processes regarding MNO data.

9. It was also concluded in 2020 that:

- A legal framework to obtain MNO-data by Statistics Sweden is lacking.
- Processing of MNO-data requires specific knowledge about the MNOs infrastructure including disturbances and upgrades of the antenna network.
- MNOs in Sweden, or at least Telia and Tre, have IT infrastructure and knowledge to process the billions of data points generated daily by their subscribers of mobile devices. Statistics Sweden does not have this infrastructure and knowledge at this moment.

10. Altogether, the conclusion was that no alternative to a partnership exists. The MNO delivers aggregated and anonymized data and related process descriptions in this partnership while Statistics Sweden's role is quality assurance by:

- Structuring the process descriptions of the MNO
- Checking plausibility
- Benchmarking anonymized and aggregated data deliveries with related statistical information
- Develop 'smart statistics' out of the data which serve general interests and are complementary to the commercial exploitation of these data by the MNOs.

11. The remainder of this paper describes how such a partnership worked in practice up until now.

IV. Process and methodology

12. Processing of MNO data is a complex process but can be divided into four main steps:

- Data-collection: collecting the raw cellular data of the devices and the related infrastructure data of the MNO
- Data-processing: (anonymization, imputing 'missing' time slots, filtering for disturbances in the network and data-deliveries etc.)
- Geolocation: relating signals of the receiving antennas to a geographical grid
- Weighting the (aggregated) sample of devices in the grid to human populations.

13. The weighting is complex because it is SIM cards (Subscriber Identity Module) that are measured. These SIM cards should first be related to mobile devices because parts of the SIM cards are placed in stationary devices. Then the mobile devices of the operator(s) should be related to humans. It should be noted that challenges in relating mobile devices to human populations remain, even if SCB manages to get data from all MNOs. On average, there are 40% more SIM cards in Sweden than persons. The complexity of relating SIM cards, via mobile devices to human populations is enhanced due to the rigid interpretation of GDPR by MNOs. MNOs do not use client data in their processes and devices are 1) anonymized at the first stage of the process and 2) followed for a limited time (24 hours) only. Therefore, the weight model is basically based on a pragmatic assumption after correcting the SIM cards for stationary devices. This assumption is that all mobile devices are at home during the night on weekdays during 'normal' working periods. Consequently, the average daily sum of devices (=baseline) during this period can be benchmarked with the registered-based population statistics at the municipality and district levels. By doing this, a weight factor can be derived. This weight factor is applied every 24-hour period that mobile devices are followed, with devices getting the weight factor of the municipality at which they were located during the night. This procedure is repeated until the updates of the weight factors during the next baseline period.

14. In 2020, it was observed that an unstable baseline of the weighting is the main cause of instability in the time series of MNO data. This instability was visible by 1) unexplained drops in estimated populations, 2) level shifts in time times, and 3) inconsistency across geographical levels (regional level vs municipality level). After investigations, it appeared that the baseline appeared to be unstable due to

- Artificial causes such as changes in the network and subtle changes in client share of the MNO at the district or municipality level.
- Real changes such as differences in human night populations during weekdays, weekends and holiday periods. An important conclusion of this investigation was that the MNO data reveal that population movements are more dynamic than initially assumed due to flex- and part-time working, multi-location living, flexible holiday periods at the individual level, etc.

15. The conclusion was in 2020: These processed MNO-data have sufficient quality to detect short-term trends in population movements due to holidays and the outbreak of the COVID-19 pandemic. These processed MNO data, however, are not suitable for official statistics yet because the quality is insufficient for level estimates and measuring long-term (year-to-year) changes precisely.

16. Even if Statistics Sweden manages to get data from all MNOs, the challenges in weighting remain. Sweden has more devices than humans: on average, there are 1.4 mobile phone subscriptions per person in Sweden.

V. Collaboration between Statistics Sweden and Mobile Network Operators (MNOs): the knowledge aspect

17. To resolve the issues mentioned in the previous chapter, Statistics Sweden has assisted Telia by structuring existing documentation about their processing of MNO-data into a standard reporting standard for official statistics: SIMS (Single Integrated Metadata Structure). The SIMS report can be considered a win-win situation for Telia and Statistics Sweden; Telia has structured information about their processes related to quality. Statistics

Sweden has the required minimum information about the processing of MNO data. We are currently working on a SIMS report for the processes of the other operator (Tre), expecting to have the same win-win situation.

18. The role of Statistics Sweden in the descriptions was and is passive and facilitating. The reasoning behind this position is that Statistics Sweden does not have superior methodological and IT expertise in these processing parts to improve the existing processing systems of the MNOs considerably at short notice.

19. Statistics Sweden's contribution to the collaboration has been active, however, regarding the weighting part. This is due to its complexity and the common observation in 2020 that Statistics Sweden has more experience and knowledge in relating observations to populations than MNOs. Consequently, Statistics Sweden has proposed improvements in the weighting methodology, which are partially implemented by Telia. At the same time, a benchmark report was developed in which (monthly) estimates derived from MNO data are compared with related official statistics. Related official statistics are, among others: population statistics, registered based commuting statistics, education and income statistics. This benchmark report can be updated monthly. Its aim is, together with the process descriptions, quality management. More specifically, its aim is:

- Checking the plausibility of the MNO estimates and detecting level shifts in time series at an early stage (note that level shifts may also occur after improving the methodology)
- Detecting real changes in population movements and evaluating which 'smart statistics' can be developed out of MNO data.

20. This benchmarking, based on open statistical data, is appreciated by the MNO and may be one of the pillars of a long-term partnership.

VI. Collaboration between Statistics Sweden and Mobile Network Operators (MNOs): smart statistics

21. The MNO data are subdivided into activities and trips. An activity is: A device is at least 40 minutes within an hour at the same location. With the help of the weight model, activities can be related to population estimates in districts, municipalities and regions during nights and daytimes. A trip is a movement of the device between two activities. It can be related to travel. Statistics Sweden receives anonymized and aggregate activity and trip data at four (geographical) levels: grid, district, municipality and region. Data deliveries are from January 2019 up to last month (April 2022 at the time of writing) with a temporal resolution of one hour. Statistics Sweden receives similar data from Tre from 2021. Data deliveries are used for benchmarking and exploring the possibility of producing statistics.

22. Potential smart statistics which can be developed out of these data are monthly or weekly statistics about 1) dynamic night populations and 2) the ratio day/night populations and related travel statistics. The high population shows, among others, clear seasonal differences in population activities per municipality, region and district. They show, among others, population movement from cities and suburbs to the countryside during the summer (Annex 1). Statistics on this topic serve the general interest (for example, Road Agencies and Regional Health Care Organizations). The ratio of day/night population activities can, be related to daily commuting and show long- and short-term effects of the pandemic on daily commuting patterns (Annex 2).

VII. Next steps

23. It should, however, be noted that two challenges must be worked out more extensively before being able to publish such smart statistics:

- Benchmark methods to detect and correct for level shift should be tested more extensively.

- Marketing of these potential ‘smart’ statistics should be successful.

24. Regarding marketing, it should be noted that the interest in such new statistical products is beyond discussion, but they are not required by any regulation. Therefore, they should – at least for the next period – be positioned as a high-quality complement to the use of MNO data by MNOs themselves for commercial activities. Partnerships with MNOs may be based on this ‘complementary statistical output relationship’ in addition to knowledge exchange and benchmarking of the results.

25. A legal base to get access to these privately owned data sources is needed to get a formal ground for guaranteed access to these data. However, partnerships with MNOs remain necessary even with a legal base because Statistics Sweden needs their expertise in the infrastructure (antennas). Furthermore, apart from the legal and practical aspects, it seems not to be cost-efficient that Statistics Sweden develops its own IT infrastructure and processes adjacent to the processing system MNOs use for commercial purposes.

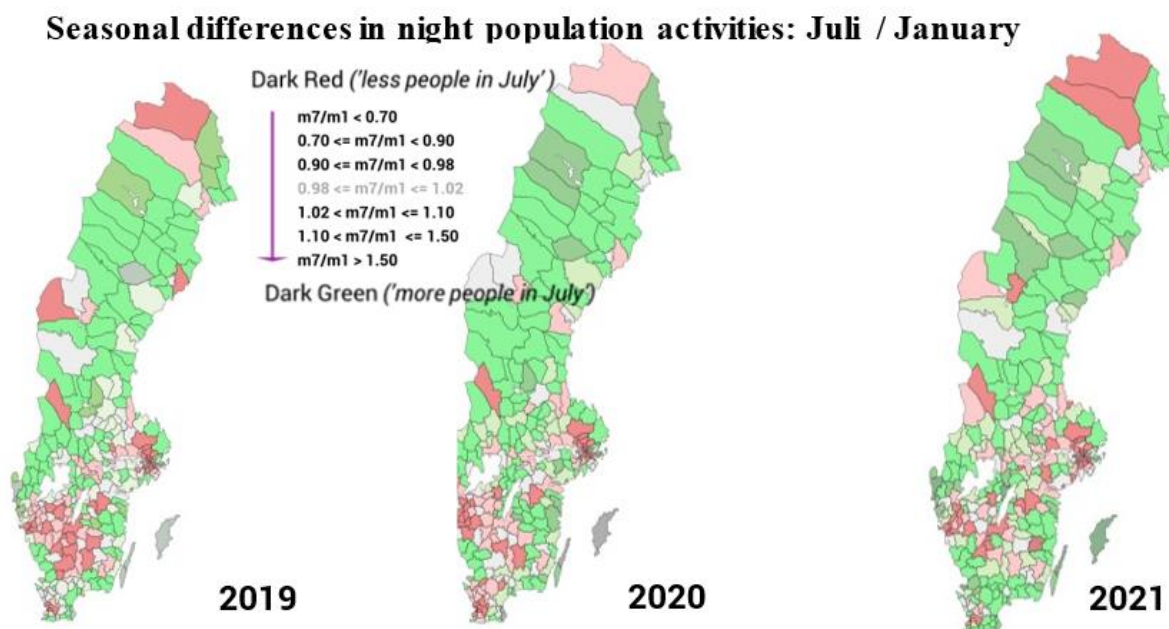
26. More international collaboration is also desirable to share knowledge and increase quality. For example, Telia’s market share is quite similar in Norway, Sweden, Finland and the Baltic countries, and Telia crowd Insight covers all of these countries. Although contacts exist with other statistical offices in the Nordics, a more intensive collaboration between statistical offices (also at the strategic level) about getting access to and analyzing data from MNOs which are active in several countries would lead to cost-efficiency and better international comparability of statistical results.

VIII. Conclusions

27. Statistics Sweden has worked with MNO-data since 2020. Without a legal basis, it should be considered as a fact that MNOs in Sweden only want to deliver aggregated and anonymized data to Statistics Sweden. They are, however, willing to share documentation if the legal aspects are covered. The lesson learned is that these legal aspects take time. After having overcome these obstacles, the collaboration gradually changed into a partnership in which the role of Statistics Sweden is facilitating in describing the processes but active in relating the weighted MNO-data to official statistics and improving benchmark techniques to monitor the quality. The aim for Statistics Sweden is to develop smart statistics about seasonal dynamic populations out of these data and statistics about commuting and trips in urban areas. For this purpose, a continuation of partnerships with MNOs is needed, based on the business model in which official statistics are positioned as a high-quality complement to commercial activities on MNO data by the MNOs themselves. A legal base is needed to have a formal ground to start such partnerships.

Annex I

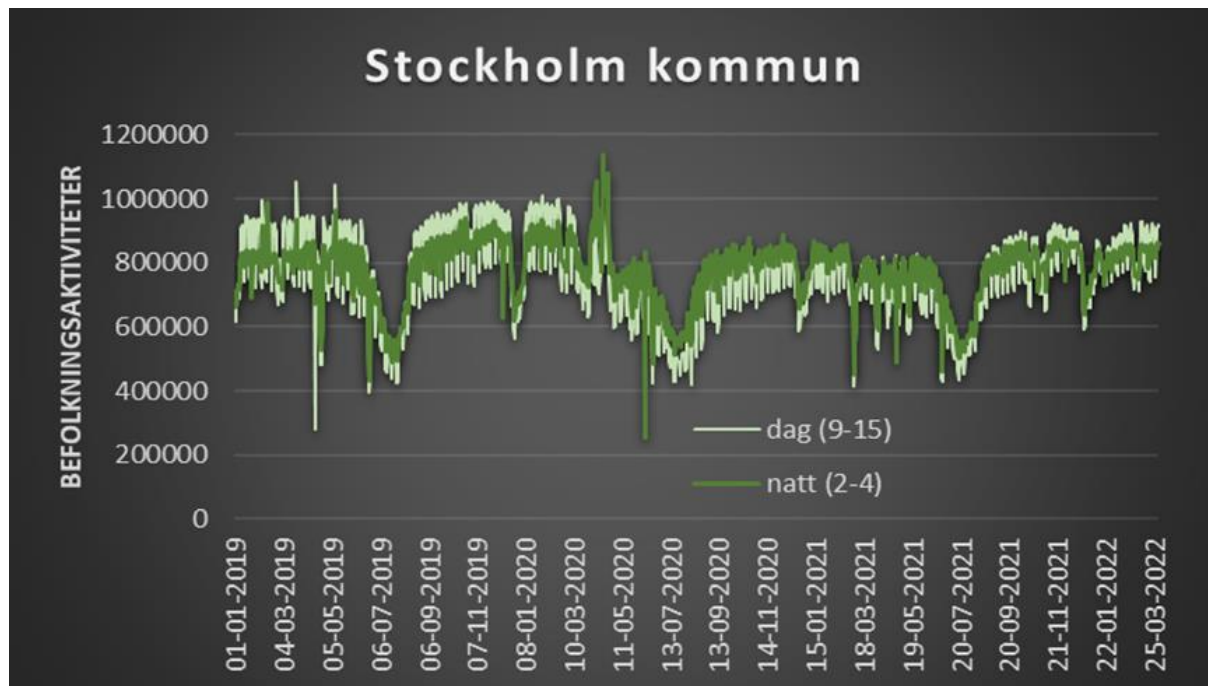
Seasonal differences in night population activities: July – January



Example of seasonal variations in night population activities. Comparison between July and January in the same year. Green areas are municipalities with higher population activities in July. These are in most cases municipalities on the coast, in touristic areas of densely populated municipalities with many 'leisure houses'. Red areas are municipalities with lower night populations in July. Lower night populations are in most cases observed in cities and suburbs or in the northernmost part of Sweden in municipalities with iron mines.

Annex II

Time series of day and night populations activities for Stockholm municipality



Time series of day and night populations activities for Stockholm municipality, the capital of Sweden. Note that before the COVID-19 pandemic the day population activities during weekdays are higher in Stockholm, together with the associated trip data suggesting commuting to Stockholm for work. These higher daily population activities during weekdays disappear during the COVID-19 pandemic and the related recommendations regarding home-working. Higher day populations return from summer 2021 but to a smaller extent than before the pandemic. Opposite patterns are found in the suburbs.

The figure also shows the lower populations during the summer and Christmas period in Stockholm (and its suburbs).