ESSnet Big Data II

Workpackage I
Mobile Network Data

Deliverable I.1 (Access)
Access to mobile network data: updated review

Final version, 08 April, 2021

Prepared by:

- M. Suarez-Castillo (INSEE, France) - S. Hadam (Destatis, Germany)
- M. Poulhes (INSEE, France) - N. Rosenski (Destatis, Germany)
- B. Loutrel (INSEE, France) - M. Offermans (CBS, The Netherlands)
- E. Coudin (INSEE, France) - M. Tennekes (CBS, The Netherlands)
- R. Radini (ISTAT, Italy) - T. Tuoto (ISTAT, Italy)
- D. Salgado (INE, Spain)

Workpackage Leader:

David Salgado (INE, Spain)
david.salgado.fernandez@ine.es
telephone   : +34 91 5813151
mobile phone   : N/A

Contents

1

Introduction

This document focuses on diverse aspects related to the access to data for a sustainable production of official statistics in multiple statistical domains. The access to mobile network data by National Statistics Institutes (NSIs) is an intricate complex issue already tackled during the ESSnet on Big Data I (WP5.1, 2016; WP5.2, 2017). It is also the focus of another working fora, such as the ESS Task Force on Big Data/Trusted Smart Statistics and its internal Working Group on Privately Held Data or the UN Global Working Group on Big Data.

In the documentation of the ESSnet on Big Data I, we already shared the situation about the access to this data source in the ESS (WP5.1, 2016), the results of an international workshop with European NSIs, Mobile Network Operators (MNOs), and other international institutions, and a preliminary analysis about the different entangled factors affecting the (lack of) access to data (WP5.2, 2017). The situation has not been unravelled yet and more work is needed to incorporate this source in the production of official statistics. In an attempt to advance in this status quo, this document concentrates on two concrete contributions seeking further steps to build trust, to identify critical issues, and to agree on a course of action.

The first contribution goes beyond our former analysis about the access to mobile network data and collects the feedback from different European MNOs on the different related factors. If the ESS is willing to make progress in the use of this data source, we argue that it is fundamental to collect and analyse the viewpoints from data holders (MNOs in our case). In this line, we prepared some general guidelines to hold a bilateral dialogue with each MNO aiming at understanding their concerns, their objections, and their suggestions for a future collaboration. We have gathered the feedback from several MNOs in the countries of the partners of this Work Package. This is the content of Part I of the present document.

On the other hand, also with the objective of making progress, we envisage potential collaboration scenarios. The different factors mentioned above and their interrelation give rise to a complex situation which requires novel formulas of collaboration. We want to share our view about how all this situation is indeed the materialization of the contents of the Bucharest Memorandum, that is, a new age for Official Statistics bringing, among other things, new ways of collaborations with data holders and new roles for statistical offices. This is the content of Part II of the present document.

It is important to underline that this document has been elaborated during the long-standing outbreak of the COVID-19 in Europe, where MNOs have closely collaborated with European and national institutions to fight against the virus. The collaboration collected in this document represents an extra effort which is publicly acknowledged.

Part I

Feedback from Mobile Network Operators

Guidelines for a dialogue

We describe below the structure of the guidelines to conduct a dialogue with the MNOs in each country. The ultimate objective is to collect their feedback as faithfully as possible about the access to mobile network data for the production of official statistics and about potential collaborations with NSIs in this production. Based on our preliminary analysis during the ESSnet on Big Data I (WP5.1, 2016) these guidelines show the following structure:

- General Introduction.

- Partnerships.

- Insights and statistical outputs.

- Business interests.

- Legal issues.

In the following sections we provide a short description of each of these parts.

2.1. General Introduction

The introduction to the guidelines are clearly intended to describe the context for this dialogue with MNOs. In this sense, different international initiatives are shortly mentioned with an emphasis on those by the ESS, especially the antecedents by the ESSnet on Big Data I and the ongoing efforts by the ESSnet on Big Data II.

After stating the complexity of the access to this data source for MNOs, involving different departments, units, and profiles in these companies, we identify as our communication targets those officers within each company both having a knowledge of the issues underlying each part and/or having a (partial) responsibility on the decision to share data and work together with NSIs. MNOs are indeed complex organizations and the decision and strategy to share data and set up collaborations involve different units in the companies.

The guidelines are intended to provide a core structure to establish a conversation between NSIs and MNOs. The ultimate goal is to build long-established partnerships to set up a joint production framework. To this end, it is advisable that NSIs are timely acquainted with MNOs' perspectives on the different issues included in the guidelines. All questions described below are directed to MNOs so that NSIs can be aware of their viewpoints.

In the dialogue, a clear statement about the treatment of the information provided by MNOs is included, communicating that it will be the core of one of the deliverables of the ESSnet on Big Data II, released in due date and form through the official dissemination channel of the project. Anonymity regarding the input provided by each company will be respected, especially if explicitly requested by MNOs (examples included in the next section). Again, the priority is to set up an honest, direct, and fluent dialogue with MNOs so that their standpoints are clearly identified and acknowledged by NSIs, Eurostat, and the ESS in general. No intention to reveal or undisclose sensitive information is pursued.

A very short explanation of the structure of the dialogue is explained prompting for a wide and open communication going away from rigid questionnaires, closed answers, and pre-fixed questions, since every situation with every MNO and every country is different. Indeed, the wording is not essential, but just the means to achieve this and this is left to the judgement of statistical officers to conduct the dialogue in the more efficient way. Every single aspect in the guidelines is preceded by a short description of the context motivating the inclusion of this aspect in the dialogue.

Every NSI has absolute freedom to make an adaptation of these guidelines not only to the specific country but also to the specific MNO. Indeed, should any issue that the company considers relevant be missing, both the NSI or the MNO enjoys absolute freedom to include it at the best convenience.

2.2.   Partnerships

Partnerships between MNOs and NSIs were already identified in the ESSnet on Big Data I as the optimal solution to incorporate mobile network data in the production of official statistics. They are still the preferred goal in the current project. Therefore, we want to identify those common points of interest in the production of statistics between MNOs and NSIs and those elements in NSIs which MNOs perceive as attractive enough as to build a collaboration around them. In this objective of building a "win-win collaboration" we include the following questions:

Q1 How do you see the contribution to a collaboration with official statistics for the operator?
This question is devoted to get general feedback about the collaboration with Official Statistics and to identify both beneficial and critical issues for the MNO.

Q2 Would you be willing to enter into a research partnership with NSI [use your case]? If not, why? If conditionally, what would these conditions be?
This question is devoted to get feedback about the feasibility and consideration of setting up a research partnership between an MNO and an NSI. In the negative case, we need to understand those issues preventing us from setting up such a collaboration. In the positive case, we need to understand what restrictions are to be met.

Q3 Would you consider improving the quality of statistics produced by MNO [use your case] using additional data from NSI [use your case] an asset for the operator?
This question is devoted to acquaint the consideration about the potential improvement of statistical outputs based on mobile network data by the MNOs by possibly integrating additional data from NSIs.

Q4 If a common methodology is developed for the production of statistics of public interest, would open source software be an impediment? What obstacles do you identify?

This question is devoted to identify potential issues in the development and usage of open source software in the potential production of statistics of public interest. The statistical methodology to produce official statistics must be open according to quality standards. We need to identify potential obstacles for this in a possible collaboration.

2.3.   Insights and statistical outputs

A priori concern in the production of statistical outputs and insights from mobile network data may arise due to a potential collision or intersection of scopes for these statistical products by the private and the public sectors, thus negatively impinging on setting up collaborations. In this line of thought, we need to identify what kind of statistical products are of primary interest for MNOs to analyse potential alignments with the public interest. Therefore, the statistical products produced by the NSI should be clearly identified, in the interest of end users and carefully with respect of MNOs business interest. We formulate the following questions:

Q1 Mobile network data is used to produce statistics of private and public interest. Could you identify what, in your opinion, is the main added value of mobile network data compared to other available sources of information (surveys, administrative data, population census, etc.)?
This question is devoted to identify the added value of these data according to MNOs both for private and public purposes especially when compared with other traditional data sources. The identification of this value as perceived by MNOs will help NSIs to approach partnerships in a more adequate way (e.g. reinforcing this value integrating more data or with improved methodology or with quality indicators).

Q2 What are the different types of target customers of your statistical product portfolio? For these customers, can you outline the statistics provided/demanded? Do you know their use (operational management? Analysis? Risk measurement?)?
This question is devoted to understand the kind of statistical outputs demanded by their customers. In this understanding a more adequate collaboration can be possibly constructed. Knowledge about the use of these insights is valuable for NSIs to understand what the public sector can contribute for the general interest of society.

2.4.   Business interests

This part is devoted to identify and understand the business interests of MNOs when producing and commercializing these statistical products. Again, a clear understanding of these details will help NSIs identify potential obstacles and enablers to design and implement feasible partnerships.

Q1 Does MNO [use case] develop an economic activity around the production and commercialization of statistical products based on mobile network data?
This question is devoted to understand if the responding MNO produces or not statistical products.

A. In the positive case:

Q2a Could you briefly describe the history of your business and commercial activity around the production of statistics from mobile network data? [Suggested items: Date of creation, international development, turnover, growth prospects, risks, periods of weakness, etc.]
This question is devoted to gain insight on the context of the MNO and its experience

in producing statistical outputs. This information is both directly and indirectly valuable for NSIs because it also provides insights about the expert vision on this newly born market.

Q3a Would you be ready to initiate consultations on the statistics that might be produced jointly with NSI [use your case] in order to find a balance between your commercial interests and the production of official statistics?
This question is devoted to get knowledge about their readiness to setting up formal contacts with the NSIs to search and find a balance between private and public interests. This knowledge is key to propose a working plan.

B. In the negative case:

Q2b Could you briefly describe the reasons leading to this decision?
In some countries, MNOs are not willing to enter into the statistical business. Understanding their position and the reasons behind this choice is necessary to reach an equal treatment for all data holders regarding mobile network data.

2.5. Legal issues

Beyond doubt, legal issues are perceived as the prickliest concern in getting access to data for the production of official statistics. At least three different perspectives enter into play in the legal considerations, namely, privacy and confidentiality of data (with the corresponding European and national regulations), telco regulations, and statistical legislation. Enough clarity is not interpreted in the different Laws, thus giving rise to risk perception hampering collaboration. In this context, we pose the following questions:

Q1 According to the e-privacy directive [and its counterpart in the national law, if applicable], the use of the postal code (or similar) and electronic communications is restricted regarding the usage of mobile network data. Has this regulation limited or is this regulation limiting your activities of the operator in the field of statistical production? This question is devoted to understand how the e-privacy directive and national translations are restricting the usage of mobile network data to produce statistical outputs and in what sense this restriction is taking place. Although private data holders and producers do not in general hold the same position in the Law with respect to NSIs (compelled and supported by legal mandates to collect and process data), it is important to understand those limitations posed by the Law on MNOs in order to build fruitful partnerships.

Q2 Regulation No. 2016/679, known as the general data protection regulation (GDPR), was released in 2016 and strictly regulates the use of personal data, conditioning it in particular on the consent of citizens. Does this regulation constitute a burden or even an impediment on the economic activity of the business unit monetising mobile network data? This question is similar to the preceding one, but focused on the GDPR.

Q3 Are the anonymisation rules enacted by the national Data Protection Authority [DPA; use the concrete name in each national case] a burden for the operator? Are anonymisation processes expensive? Has the operator already been audited by the national DPA?
The question is devoted to understand the relationship with the corresponding national DPA in each case. In particular, in their legal role to protect privacy and confidentiality, these agencies usually enact anonymisation rules which naturally impinge on the statistical production process. Understanding these details, especially regarding the processing costs, is a valuable piece of information in order to design and implement potential partnerships.

3

Feedback

3.1. The Netherlands

The current situation of access to mobile network data in The Netherlands prevents us from having a direct feedback from Dutch MNOs using the guidelines included in the appendix A. However, there exists experience in previous contacts so that we can include the views from the present authors. These views do not necessarily reflect the policies of Statistics Netherlands.

3.1.1. Data access

Access to mobile network data is –just like in most other countries- a challenge in the Netherlands. Dutch mobile network operators are responsible to ensure privacy for their customers and they take this task very serious. Moreover, operators consider mobile phone network data as part of their strategic assets, similar to their physical network, their brand name, etc.

Nevertheless, some operators feel the responsibility to allow the use of network data for the benefit of society. Mobile network data enable a very high-frequent (even up to hourly) population census. Usage for statistics and scientific research in fields like disaster response, health, mobility and tourism has already shown its potential. The wide range of application fields opens the opportunity to create a win-win business case for the public good on the one hand, and for commercial benefit on the other hand.

Mobile network data use and in particular commercial use is a very sensitive topic, both from a legal and perception perspective. The laws on mobile phone traffic, created already decades ago, are rather strict and basically prohibit almost any form of data reuse, except for internal operator purposes to optimize their network. The main goal of these conditions was to ensure privacy. This was many years before GDPR entered into force, and also before big tech companies became very active in reusing customer data.

We have to take into account that mobile network data, clearly, are extremely privacy sensitive and that citizens should be protected against misuse such as illegal surveillance by companies or government bodies. A less obvious issue is that we have to avoid the risk that any accidental identification of individuals can occur out of aggregated anonymised data. This risk is higher than it may seem at first. New studies like those by Tu *et al.* (2018); Zang and Bolot (2011); de Montjoye *et al.* (2013) show that information on individuals can be recovered from anonymized origin-destination tables in certain cases, given sufficient computing power and some auxiliary information. To mitigate this risk, advanced statistical disclosure control techniques should be applied when creating origin-destination tables. These techniques should

be kept state-of-the-art, and updated as needed, before publication of this type of tabular information.

The following aspects can be distinguished that come into play when talking about access to mobile network data:

1. The legal framework, based on GDPR and legacy (supra)national telecom laws that have been set up to protect fundamental rights of citizens.

2. Ethical considerations beyond the legal framework.

3. The operators that store the data generated by their networks and their customers.

4. Government bodies at (supra)national, regional and local level, statistical offices, that are interested in information (and ask for it) for operational, planning and strategic purposes in several fields.

5. Private companies (retail trade, hospitality services, . . .) that may want to use these data to optimize their services for customers.

6. Public opinion and the media.

Below we discuss each of these aspects briefly.

1. The legal framework based on GDPR is up to date and in principle works out fine, but it has to be considered together with specific telecom laws that can be decades old. The relation between such laws may cause a lot of discussion between legal experts about which law takes priority, and what is allowed in each case and what not. This situation applies to anyone who handles mobile network data and therefore the legal framework is an important and nontrivial aspect to consider when accessing mobile network data. The new e-privacy directive hopefully solves this difficult situation. Important requirements of these laws are that data should be used only for aggregated anonymised results and data storage should be minimised to the extent possible to protect civil rights and interests of customers.

2. The legal framework is focused on protection of the individual. As a consequence, profiling of individuals is of course not allowed. Modern machine learning (or artificial intelligence) methods, however, allow detection of important new phenomena. Although such methods could become important in an information based society, they might cause unwanted side effects for certain groups of people. For example, identification of the behaviour of specific groups, detected by machine learning classification algorithms, may lead to stigmatisation. To avoid that, ethical considerations should always be part and parcel of the analyses. It is not easy to design laws or norms for this type of situation because the number of variations is unlimited; in any case, at present such laws and norms do not exist. An important advice to deal with this lack of legal guidance is that algorithms should be made open/public for reviewers to allow a form of democratic control or review.

3. Operators are continuously faced with new challenges to protect the data generated by their networks and their customers. Security constantly has to be upgraded to protect network data and the networks themselves from hackers, or even attacks from certain states. Introduction of 5G further expands the size and frequency of data that has to be stored to keep the network up and running. Therefore it is important to minimize data traffic as much as possible to reduce risks. Data should be stored only on the premises of

the operator. This is both efficient and the best guarantee that customer data is safe. In addition, analytical results need to be created in a dedicated process to ensure anonymity of mobile phone users.

4. Policymakers and other potential users in governments face challenges relating to (hyper)mobility of the population that have implications for, e.g., safety, health and infrastructure. This hypermobility appears on different spatial scales. On the one hand, it does not stop at national borders and information about it may yield insights in cross-border phenomena. On the other hand, information about local mobility can also be very relevant. A city, for example may want to know how many people visit it during the day for commuting, shopping or leisure ('daytime population'). And for every event –no matter how big- it is interesting to know how many people attend. A normal census or registration only provides static information on the population- in fact, where people sleep, not where they are active. Moreover, the rise of online bookings via platforms changed the whole tourism industry, making it rather difficult to compile reliable tourism statistics based on hotel registrations only. To stay relevant for society, NSIs would need access to mobile network data (of course within the limits given by GDPR and other applicable legislation).

5. There is a growing demand for geolocation-based data for commercial purposes, which causes a certain pressure. This demand should be taken seriously and it may even interfere with use by governments. It presents an opportunity for operators and other private companies to create financial value by analyzing data. Shops may want to know, for example, where their customers are coming from or how much time they spend shopping. This a new and expanding field and it is difficult to see how it relates to current legislation and civil rights. Big tech companies already have pilots in place to provide statistical results for the private sector and the operators no doubt will investigate if they can enter this profitable data analysis market. The expectation is that this market will only grow over time.

6. For all stakeholders, public opinion is important. Here NSIs can play an important role in informing public opinion in a balanced way. They should explain which privacy (and other) risks may occur and how to manage them properly. It is even more important to explain in a transparent way how data is being processed and what are the benefits for society. To strengthen the NSIs societal role, new research on how to improve privacy is needed.

3.1.2. Challenges

Given the above comments, the key challenges concerning access to mobile network data can be summarized as follows:

1. **Outdated (legacy) telecom** laws that give room for interpretation on what is (not) allowed with reusing mobile network data.

2. **Perceived and real re-identification risks** following the development of new and complex methods to recover data on individuals from aggregates. Anonymisation and data protection can still be guaranteed in practical cases, but advanced disclosure protection measures have to be applied and new potential risks have to be monitored. This requires awareness, dedicated research and specialized skills to avoid lagging behind and incurring unnecessary risks.

3. Privacy authorities have some **national freedom to interpret European rules** and recommendations. They can start investigations given point 1 and 2 to show they are actively

protecting the rights of citizens. At the same time, there expertise is rather in the legal than in the technical/methodological domain; and the topic of mobile network data use may generate highly emotional sentiments. This complicates rational discussions, consumes time and creates room for misunderstandings because the whole topic is **complex from a legal, societal and technical perspective**.

4. It is essential to explain why access to these data for statistical purposes is beneficial from a societal perspective and which safeguards are in place to protect privacy. As statistical agencies, we tend to overlook that because we from our perspective and the experience we have built up, we are already convinced. An open dialogue with and transparency to the general public is therefore important. Apart from statisticians, policy makers and other users should join in the communication to present their position. Good and effective communication, however is far from easy.

3.1.3. Needs for the future

What is needed? In order to deal with the three key challenges identified above, we envisage the following road ahead:

1. For statistical and scientific purposes, the forthcoming EU e-privacy directive should foster **device-level data processing at the earliest steps of the process incorporating anonymisation and privacy by design from the onset**.

2. Minimize **data traffic**. Data should be stored and processed as much as possible on-premise at the operator. Only well-protected aggregated data should be shared with statistical offices and academia. We should realize that an alternative world with commercial data brokers and big-tech companies having access to these data is undesirable with uncontrollable and in transparent impact on privacy and other civic rights.

3. More attention for and research on **statistical disclosure control**. Statistical disclosure control is never finished. New studies that provide methods to hack **aggregated** data should be investigated and measures should be taken in the data processing to guarantee anonymity of the output. Awareness of these potential risks and protection against it should be increased, similar to the efforts to improve cybersecurity.

4. Develop a communication strategy that addresses not only privacy, but also the societal benefits of making statistical outcomes available.

3.1.4. Conclusion

Despite all the research and evidence to responsibly create statistics with MNO data, all research has been stopped. The method and technical details are largely ready. This also applies to the privacy assessments. The statements in the press by various parties about alleged privacy violations make it impossible to do any more research with MNO data because of public image risks. The whole situation is thus stuck. From an initial frontrunner in research, the Netherlands is now one of the few countries in Europe that is not using this data for the COVID-19 crisis. It is unclear if and when this will change.

3.2.   France

3.2.1.   Introduction

In France, there are four Mobile Network Operators (MNOs). Three of the MNOs have developed in recent years statistical products for sale. The first private statistical solutions were launched in 2013 or 2014. These commercial offers are widely known by local public authorities and in the transportation sector.

The French National Council for Statistical Information (CNIS) qualified, in its 2019-2023 medium-term outlook, the present population statistics as information of public interest, complementary to residential population and encouraged experimental investigations on new data sources (`https://www.cnis.fr/instances/moyen-terme/`). The French Statistical Institute has investigated since 2016 how mobile network data can be used to produce structural present population statistics, appropriate to calibrate local public services (care supply, sanitary checks, firefighter services...).

A first round of exchanges between INSEE and MNOs was conducted in 2015 (Rapport Bon) (Gregoir and Dupont, 2016) when the Institute presented to the MNOs its needs and framed the principles for the co-existence of statistics of public interest and for-profit statistics. At this stage and since then, MNOs did not volunteer.

In the context of the ESSnet Big Data II and the WPI deliverable on Access to MNO data, a second round of exchanges were conducted based on bilateral interviews with MNO representatives so as to better identify the obstacles in implementing a *mutually beneficial* partnership. MNOs were asked to comment and react in an open discussion to a set of questions (see Appendix) based on four main topics: (1) a partnership scenario, identifying their potential benefits (2) value segments (type of offers and clients) (3) legal aspects, and (4) business model.

All the four MNOs agreed to have this discussion. It was agreed that the collected information would not be directly identifiable. Interviews were conducted by phone or in-person and lasted between one and two hours. A couple of questions were sent beforehand to frame the discussion.

We first lay out how the operators described their data activity, its perspectives and associated risks. The possibility of a successful partnership depends indeed strongly on the MNOs' ability to process, collect and store mobile network data for their own business. Second, we report their main reactions on the offered partnership.

3.2.2.   On the activity itself

The business units producing statistical outputs have mainly four types of clients: the transportation sector (e.g. bus company), the retail sector (e.g. shopping malls), the local authorities (e.g. municipalities) and the local actors (e.g. tourism development agencies). Two MNOs have specialized to better answer to particular types of clients while one MNO has developed an industrial solution (reaching a large number of clients): there seems to be a customization trade-off in the development of the activity.

3.2.2.1.   The market is still not mature

During the interviews, two MNOs qualified the market as disappointing compared to what was expected when the activity started. All MNOs insisted on the costs of investments. The operating costs are considered high in terms of information system, treatment platform, skills, data pre- and post-treatments to get the statistics consistent. Yet, if primary interest is generally high for many actors or businesses, willingness-to-pay for the service is not as frequent or rather low. All MNOs pointed out to a low profitability activity, seen as a strategic investment or waiting for the demand to develop. For this reason, some MNOs expressed concerns on the continuity of the activity. To limit the costs, one MNO externalized large segments of its statistical activity (e.g. data science, programming).

Increasing spatial precision was identified as an opportunity to grow. One MNO emphasized that one current threat to their market is the GAFA's ability to disseminate open-data indicators (rather than addressing directly the same clients).

3.2.2.2.   The legal aspects are uncertain

At the beginning of their activities, all MNOs consulted the Data Protection Authority (DPA). They presented their technical solutions and how they intended to preserve the privacy of their users. In practice, three very distinct solutions co-exist. Some MNOs keep pseudonymised data in secured environment while others perform on-the-fly anonymisation. They do not dispose of a formal agreement which would secure the activity. In particular, some MNOs underlined that each new methodological improvement entails a risk as it should be declared to the DPA, endangering the current equilibrium. It constraints evolution in the method all-the-more than the privacy protection scheme is embedded deeply in the statistical process. The current legal framework has been subject to different interpretations in practice and is currently relatively uncertain. Some MNOs fear also for their image if these treatments are interpreted by their customers as a misuse of personal data.

3.2.3.   On partnerships with NSIs

3.2.3.1.   The potential pros

- Quality label from NSI
Some MNOs expressed interest in a form of quality labeling by the NSI – for instance by communicating on a methodological collaboration. However, they warned at the same time against the risk of distortion of competition if only some MNOs are labeled by the public statistics in the process. This would create a market distortion penalizing some MNOs.

Another MNO did not express interest in quality labeling would it be granted to all national MNOs. This MNO sees the labeling as a competitive advantage; attributing the same quality label to all MNOs prevents its potential benefit. In addition, for this MNO, the quality of the statistical products is not anymore discussed with its long-term clients, notably because relevant external validation sources do not exist. In this context, this MNO suggested that NSI should frame quality standards as the ability of the statistics produced by MNOs to match some reference external data sources. The NSI would act as an observatory instead of getting involved in the method applied to network data at the lower level and instead of asking for full transparency.

- NSI as a trusted third-party
Two MNOs expressed interest for the inclusion of the NSI as a trusted third party to deal with specific questions where the multi-MNO view is key given the data (e.g. international tourism).

However, they warned that multi-MNO discussions are potentially exposing the benefit of their own investments and that such partnership is not welcomed if it would lead to a standardization in terms of privacy models (as different models coexist and have been costly to implement).

- Quality improvement
One MNO expressed interest in working transparently with INSEE on improving the methodology, while other MNOs were more reluctant to openly discuss their methods – invoking business secrets and network private and complex knowledge. One MNO expressed concerns that high-quality and highly-disaggregated validation sources are really rare thus the validation task is not feasible in practice. Another MNO declared that the demand for quality was relatively low as it only marginally deterred clients in the past, which did not aligned with some other interviewees' opinion. More or less flexible/industrialized production lines seemed to play a role as well in MNOs willingness to invest in new methods. One MNO considered that methodological improvements would be interesting only if it conferred a competitive advantage relative to other MNOs.

Experience feedback from the use of statistics from commercial offers by INSEE

At the start of the COVID-19 crisis, three MNOs engaged limited-in time collaborations with the NSI and accepted to transfer aggregated data to the NSI for free to document present population variations during lockdown and mobility indicators. Two MNOs provided statistics coming directly from their commercial offers while being reluctant to fully open their methodology and kept some parts of that methodology undisclosed for business secrets reasons. The commercial offers only partially corresponded to the needs of the NSI but were quickly available. In the crisis situation, facing some major phenomena to measure, the cross-MNOs analysis resulted in several NSI publications (Galiana et al., 2020; INSEE, 2020a,b,c). However, comparisons with INSEE reference/benchmark statistics showed also the shortcomings of the indicators based on network signaling data as produced by MNOs relative to Official statistics needs and standards - uncertainty, variability of this kind of data, and the population-adjustment process. Articulating transparency in dissemination and MNO business secrets requirements turned out to be challenging. This highlighted the limitations of relying on commercial offers – even when sharing the cost was not discussed.

It was not the first time that public statistical entities were using commercial offers of MNOs. These entities had already identified some limits: partly non-transparent methodology, difficulty to use in the long run because of break in series.

A joint work in order to improve the quality of the final statistics seems unavoidable for meeting official statistics quality standards.

3.2.3.2.   The cons

- Investment financing
Starting a meaningful partnership with the French NSI was seen by most MNOs as a prospective investment and was not rejected per se. However, some MNOs pointed out that this partnership would require substantial endeavor, but based on the current state of the market, no investment can be made today or only under strict return-on-investment conditions or external financing. One MNO insisted that a contractual commercial relationship was the only form of collabora-

tion that they could foresee. Access for free was excluded because of the operating cost and investment that the activity represented. The INE commercial relationship with an MNO was mentioned as a possible model (see INE, 2019a,b,c, 2020a,b,c,d).

- Infrastructure costs

All MNOs alerted on the fact that the (highest-value signaling) data require strong expertise on telecommunication networks for collecting and preprocessing the data. Some pointed that the Big Data ecosystem entails high maintenance costs that should not be minimized.

- Risk of the freemium offer for business

One MNO has already considered the solution of a freemium offer without putting it into practice however. For this MNO, the potential overlap between MNOs and official statistics appears minimal, and the labeling provided by public statistics could increase the demand for this type of indicators. At the other end of the spectrum, another MNO is really skeptical about the possibility of reaching an agreement on a potential freemium offer. This MNO is particularly reluctant to the idea of the publication of structural indicators by official statistics. This MNO expressed the fear that this could give rise to exploitation of these statistics by data marketing companies, cutting the market for the operator's statistics as a result. The scope of the disseminated public statistics (geographical level, time granularity) could turn out to be a key point in this discussion.

3.2.3.3.   Conclusion

Overall, MNOs did not close the door to continue the discussion on collaboration. Several MNOs pointed out that they were already using publicly available INSEE statistics in their production to calibrate their data. However, identified challenges seem too high to overcome without a strong case of the public interest for an (experimental) official statistics. Moreover, to be viable with MNOs business models, the public scope would likely be reduced – hampering the ability of the public statistics to match (all) local actors expectations. The balance and co-existence between public and private interests has still to be addressed. Defining precisely the target of the public statistics could help in assessing the market-related risks for MNOs and the needed investments and costs. MNOs were overall not unfavorable to start a concrete collaboration project.

3.3.   Germany

3.3.1.   Introduction

In Germany, there are three mobile network operators (MNOs), Deutsche Telekom AG, Vodafone GmbH and Telefónica Deutschland Holding AG, with a respective market share of one-third each in the German mobile communications market (see Federal Network Agency[1]). As there is no legal access to privately held data in Germany, Destatis currently has to rely on cooperation agreements with the MNOs in order to examine the data source and the specific purposes for official statistics. Of the three MNOs in Germany, data from two MNOs in Germany (Deutsche Telekom and Telefónica Deutschland) are currently available for feasibility studies, but only in exchange for monetary compensation. The Federal Statistical Office (Destatis) is currently receiving mobile network data from Teralytics AG, which generates and evaluates

---

[1]`https://www.bundesnetzagentur.de/DE/Sachgebiete/Telekommunikation/Unternehmen_ Institutionen/Marktbeobachtung/Deutschland/Mobilfunkteilnehmer/Mobilfunkteilnehmer_ node.html`

data from the Telefónica Deutschland network.

In order to get an insight into whether and under which conditions the MNOs Deutsche Telekom AG and Vodafone GmbH in Germany would like to cooperate with Destatis or enable data access, Destatis was able to conduct an open exchange with T-Systems, a wholly owned subsidiary of Deutsche Telekom. In particular, questions about possible forms of cooperation, data access, business interests and legal aspects were discussed. The aim was to identify which strategic, legal or technical reasons would speak for or against a cooperation and data access at the respective MNO. It was not possible to arrange an interview with the MNO Vodafone GmbH that is neither using their data for analysis nor willing to cooperate with official statistics or any other partner as far as we know.

3.3.2.   Results of the interview with T-Systems

T-Systems evaluates mobile network data from the Deutsche Telekom network. This data was processed and provided by Motionlogic GmbH, a subsidiary of T-Systems, until the end of May 2020. Due to the dissolution of the company, the business segment of data preparation for commercial purposes at T-Systems, hereinafter referred to as MNO for the sake of simplicity, was initially shut down. Note that the MNO is Deutsche Telekom and the subsidiary T-Systems is responsible for marketing the mobile network data, among other things. Currently, T-Systems is rebuilding the business segment.

3.3.3.   Opportunities for cooperation

Tiherefore, the MNO is interested in renewing the cooperation with Destatis and in setting up a new cooperation agreement. This involves, on the one hand, the processing and delivery of mobile network data and, on the other hand, a collaboration on research issues and further development of the resulting statistical outputs. Such a cooperation agreement provides restricted access to mobile network data. This includes defined data sets or statistical outputs for specific applications, which must be processed and delivered by the MNO. Unrestricted, permanent access to raw data is not possible. Due to the current business development and the creation of all necessary conditions for business and commercial activities, the MNO is currently not able to develop long-term scenarios for the cooperation with Destatis. This will only be possible once all necessary legal and technical requirements have been created. In principle, the cooperation between the MNO and Destatis offers the opportunity to discuss the production, dissemination and quality of statistical products based on mobile network data. In this context, however, it should be emphasized that such a framework or cooperation agreement involves monetary compensation for the company's services.

3.3.4.   Business Interests

The MNO's business or the commercialization of mobile network data or statistical outputs is based on three concepts derived from its customers and the direct benefit of the data, which are generally of private and public interest. In addition to purely commercial purposes, these concepts include social benefits and municipal support, for example in the form of research projects, as well as the further development of certain areas of society, such as the improvement of public transport services, eMobility and the like. For all processed mobile network data and statistical outputs, the economic activity is based on the commercial marketing of these products. In order to keep the data work profitable, also regarding research-driven cooperations, a monetary compensation for the expenses and services of the MNO must be considered. The customers of the MNO are public institutions, as well as private companies. All customers have

an influence on the statistical output, since these outputs are processed according to the wishes of each customer. The steps of data processing are based on the methodology implemented at the MNO. However, the MNO is also interested in evaluating the methodology used on the basis of the methodological competence of official statistics and, if necessary, developing it further together with Destatis.

3.3.5.   Insights from mobile network data

In principle, the MNO wants to set up its data portfolio in such a way that as many data points as possible can be included in the analysis. It is planned to use both signaling data and CDR (Call Detail Records) data. This is intended to increase the overall temporal and spatial accuracy. The methods and algorithms are currently being developed by the MNO and must be approved by the national data protection authority, the Federal Commissioner for Data Protection and Freedom of Information (BfDI). The statistical outputs, which are produced by the MNO on the basis of mobile network data, are created using additional data from official statistics. For this purpose, official data such as population figures from the census are used to calibrate estimates from mobile network data with respect to different target populations. Additional data, for example from road traffic authorities or train passenger data, is also used, for example to derive the potentially used mode of transport. In general, further data exchange between Destatis and T-Systems is conceivable. So-called white label phones are additionally used as ground truth information. The movements and activities of these devices can be analyzed without anonymization process and the processing methods can be optimized based on this information. The white label phones are special mobile devices used by employees of the mobile network operator who have agreed to process their data for this specific purpose.

A central point for a promising cooperation is the possibility to use specially created methodologies for mobile network data processing. The ESSnet Big Data working group "Mobile Network Data" is developing an (open source) program to generate harmonized mobile network data according to a specially developed mathematical-statistical method for official statistics. An implementation of this program at the MNO is under consideration. For this purpose, fundamental aspects must be clarified in following discussions whether and how the program can be implemented at the MNO. In addition, the technical and economic efforts and costs for the implementation of the program at the MNO must be evaluated in order to ensure the economic profitability of the MNO. It is not clear to what extent the possibly jointly developed statistical products will be made available to the public, as the MNO focuses on its economic efficiency. However, it has been pointed out clearly that the transparency of official statistics must be guaranteed. However, the MNO is open for a cooperation in this matter.

3.3.6.   Legal issues

Data protection regulations restrict the activities of the MNO in any case, but they are considered by the MNO to be important and essential for the processing of personal data. By working closely with the BfDI, the MNO aims to develop the best possible statistical output while protecting the privacy of mobile phone users as much as possible. A close exchange with the BfDI is particularly important, since the MNO wants to develop a procedure based on the extensive data available that will enable evaluations to be output on a smaller scale, i.e. from the current 500x500m grid cells the further development is planned to 80x80m or 120x120m. Work is also being done on the temporal accuracy of the data and therefore the statistical output. Specifically, possibilities are being examined to use real-time data with a delay of 10 minutes or less. In addition, the MNO is currently negotiating with the BfDI to increase the minimum number of mobile activities that may be transmitted to Destatis from 5 to 30 mobile activities

per time unit and study area. This means an enormous loss of information for Destatis, as the spatial and temporal resolution of mobile network data for Destatis' research purposes will be negatively affected. One of the most inhibiting anonymization rules for evaluations in Germany is the 24-hour evaluation period. This is stipulated in the German Telecommunications Act. The MNO is currently in consultation with the BfDI in order to extend this evaluation period for appropriate analyses and thus improve the evaluation possibilities for specific issues. On the whole, it is obvious that the MNO attaches particular importance to compliance with data protection rules as well as the protection of the privacy of its customers.

3.3.7.   Conclusions

So far, Destatis has access to two of three MNOs in Germany. Due to the fact that there is no legal regulation for the use of MNO data in Germany, the data access on basis of the cooperation with the MNOs does not give enough insight into the data and the data processing of the MNOs, which is required by official statistics in Germany for a regular statistical production. In order to ensure the transparency of official statistics and the quality of statistical outputs, there is a need for a deeper understanding of the methods used by the MNO. Most of them affect the business model of the MNO. For this purpose, an open and at the same time protected working environment must be created in order to not jeopardize the business interests of the MNO on the one hand and to allow disclosure of the methodologies for official statistics on the other hand. Nevertheless, MNOs are willing to cooperate with Destatis in order to gain a certain reputation and to improve their products. Destatis, however, receives the opportunity to get first insights into the methods used by the MNO and to better assess the quality of the data production process. This would be facilitated and ensured especially or only by the implementation of legal regulations.

3.4.   Italy

3.4.1.   Introduction

In Italy, the main mobile telephone operators are three: Vodafone, Wind-Tre and TIM which each hold about one third of the market[2]. In 2017, Istat entered into an agreement with Wind-Tre for the supply of CDR data traffic of a province for 5 weeks. The provider took care of the anonymization and a secure transmission protocol of this data was defined. This was an important opportunity to carry out studies on this type of data (although not definable Big Data) in order to investigate problems relating to localization and estimates referable to populations usually residing in small areas. The trial is currently suspended because of the request of clarification by the Authority of the Guarantor: about the purposes of the processing, the protection measures adopted and the risk assessment for citizens, with particular attention to the probability of re-identification.

The Data Protection Impact Assessment document is currently under consideration by the authority.

This type of access received resistance from both the MNOs and the Guarantor, but if properly designed (according to the criteria of privacy by design) it is highly effective in the definition of algorithms which can then be shared or transmitted to the providers to be performed on

---

[2]`https://www.agcom.it/documents/10179/20440899/Allegato+16-10-2020/`
`a7fb0d05-7630-471b-aeef-66da1b755b7c?version=1.0.`

MNO production systems.

In 2020 Istat activated an agreement with Vodafone Italia (hereinafter VI) regarding another type of access to data, mediated by the MNO. It has established an agreement composed of two Sprints that provide a close collaboration between teams of Data Science Istat and Vodafone, in order to achieve probe data processing algorithms to realize two outputs of interest to Istat. This is a first opportunity of collaboration with Vodafone that have to be stabilized.

3.4.2.   Opportunities for cooperation

From the conversation with the provider, it emerged that they are also fully open to this collaboration, it will require a significant effort at first to agree on the scope, but are willing to make available their existing Big Data framework and the "Vodafone Analytics" product available in several European countries.

Some of their statistical product portfolio are described in https://www.vodafone.com/business/cloud-and-hosting/digital-services/vodafone-analytics.

The agreement with Vodafone provides for a duration of two years with various job opportunities in the form of Sprints.

Sprints consist of three phases: planning, implementation, review; the phases are recursive. In the planning phase, NSI together with the MNO defines: the objectives of the analysis, the field of observation, the product requirements and a first draft of the data pretreatment algorithm. In the implementation phase, the MNO performs the selection of the data, the agreed elaborations and produces a report that is used as input in the review phase, where it is possible to decide either to review the planning or to detail the processing algorithms. The review phase is performed together with the MNO but also autonomously by the NSI which evaluates the report data using other benchmark sources.

This model of collaboration has already been adopted by VI with various other public administrations, governments and municipal councils. We are discussing a collaboration model that satisfies both parties and also takes into account the fact that the work involves a cost to Vodafone, mainly for storing and processing data in the cloud and the data scientist's time.

3.4.3.   Business Interests

The mobility data analysis projects conducted by Vodafone so far have concerned mobility and in particular tourism. They were conducted on behalf of public administrations and also private entities. In most cases, the customer trusted Vodafone's analytics experiences and so bought a turnkey package. Collaborating with Istat, they expressed their interest in an exchange of skills in data processing and data analysis. They can share more detailed documentation, and many controls are part of the existing frameworks. Vodafone Italia, having an international structure, is interested in collaboration models that can also be adopted in other European countries in which it operates. In general, it is willing to share the indications of the algorithms developed for the production of outputs of interest for the official statistics and Technology stack but not the Algorithms (written e.g. in pseudocode), the developed software or Network configuration informations. Furthermore they say they are willing to further analysis and assessment for the use of the secure multiparty computation technique.

3.4.4. Quality approach

VI is open to sharing its quality management framework and detailed documentation. These approaches include: algorithms dedicated to the delete of opted-out customers, apply algorithms to correct Vodafone bias, to extrapolate results to total population. They use ground truths to fine-tune these algorithms (events, sport games, data from Business customers...). The coverage of antennas can be measured with the use of the GPS signal of the SIMs that have issued particular tracking authorizations. This technique can be used where coverage areas are more uncertain.

3.4.5. Legal issues

They allow all customers to be excluded from these analyzes. In fact, they can manage the inclusion or not of this specific authorization in MyVodafoneApp and Web self-care, even if the data is anonymous and the insights are aggregated statistics. Then they are excluded and consequently corrected during the application of the bias and extrapolation techniques. They spread only aggregated data and also on these are implemented privacy output algorithms.

3.4.6. Conclusions

The two experiences of processing mobile network data, the first conducted directly by Istat on Wind-Tre's data and the second conducted in collaboration with Vodafone, have strengthened the idea that the process of accessing and processing these data must be conducted in close contact with the MNOs.

Even if the MNOs have consolidated experiences of data analysis and complete knowledge of the raw mobile network data, they need a paradigm shift in the production of the necessary outputs for official statistics. For Istat it is essential to know some specific information that allows the construction of quality indicators and the complete transparency of the processing algorithms. This would be guaranteed if the NSI could define and simulate the processing algorithms on a portion of data provided by the MNOs, as performed by Istat on Wind-Tre data. The simulation activity could make the algorithms known to the NSI and so publishable, while the software development would remain the property of the MNOs.

This solution requires a legal regulation on mobile network data access, in particular from the point of view of privacy management. Istat has initiated a dialogue with the authority of the privacy Guarantor in order to define both the management criteria for the access to data and any additional anonymisation criteria.

The interest in our collaboration with the MNOs remains significant, not only in order to create new business opportunities, but also to have an exchange of skills that allows to build new analyses and define new quality paradigms. All this would allow to give the MNOs a quasi-certification of their products deriving from the collaboration with the NSI.

3.5.   Other countries

The rest of country members of this work package does not have access to mobile network data for the present project and no information has been possibly collected. In this context, Statistics Romania (INS) has found out that a strong connection with mother companies in this country is fundamental. Statistics Spain (INE) has reached a verbal agreement with one MNO to hire a research team with staff (Big Data architects, data scientists, software developers, project

manager), software tools, and hardware infrastructure to investigate the partial implementation of the ESS Reference Methodological Framework described elsewhere (see e.g. deliverable I.3 of the present project). Further work is ongoing to carry forward this agreement in compliance with legal regulations for public contracts. Estonia does not have access to mobile network data.

4

Conclusions

Access to mobile network data for the production of official statistics is beyond doubt an intricately complex issue with many interrelated facets. As shown in the preceding sections and in preceding projects (WP5.1, 2016; WP5.2, 2017) we have already accumulated some evidence to consolidate some conclusions. The bare truth is that as of this writing no access to mobile network data for sustainable production of official statistics by NSIs has been granted in the ESS. Even access for research purposes by NSIs also faces limited and restricting conditions.

As a first general comment, we believe that the context provided by the Bucharest Memorandum (DGINS, 2018) is highly adequate to frame our current conclusions. This document provides strategic insights and recommendations for the immediate future of the European Statistical System. More concretely, the Directors-General of NSIs and Eurostat, considering among other things that "official statistics need to **remain relevant** [our boldface] in the future and continue to satisfy users with high quality statistics", "recognise the need to further develop the legal framework at European and national level to reduce **obstacles to the access** [our boldface], use and integration of [new digital data sources]". With mobile network data, we are facing multiple issues to reduce these **obstacles to the access** to **remain relevant** and produce timely and disaggregated statistics. In this line of thought, the ESS and Official Statistics in general should consider its presence in the international community, especially, in those sectors generating and holding digital data as a direct consequence or as a by-product of their economic activity. The mobile telecommunication industry is a clear example. In our view, this is in much relation with the European and national data strategies and the creation of business-to-government data stewards in these data-holding organizations. When reading proposed profiles for these new professionals (Verhulst *et al.*, 2020), their roles and responsibilities easily remind those of a statistical officer in an NSI. The presence and preponderance of Official Statistics in these initiatives can provide an indirect indication of how much/less relevant statistical offices are currently considered in the new datafied society.

The diverse situations depicted in preceding sections allow us to propose as main conclusions the following list of reflections:

- We are progressively reaching the common opinion about the need to elaborate a more specific legislation framework regarding the access and use of mobile network data for the production of official statistics. This entails a further clarification of the legal support by European and/or national Statistics Acts so that statistical offices can request access to these data, a disambiguation about the clash with (sometimes outdated) telco regulations, and a clear statement of the limitations posed by personal data protection regulations to the production of official statistics (survey data traditionally used to produce them are

identified personal data; are digital data to receive a different legal treatment?).

Furthermore, we also detect a need for the specification of roles and responsibilities for every actor in the process, a more specific account of what data are to be accessed in the rich data ecosystem of mobile telecommunication networks and under what conditions. This must be undertaken under a clear technical and statistical understanding of these issues and its implications for the production of official statistics. At any time, at least three aspects must be jointly conjugated, namely (i) the legal support to provide a sustainable data provision for Official Statistics, and hopefully, for the Public Good beyond the European and National Statistical Systems, (ii) the legitimate business interest of MNOs to monetise their data, and (iii) the strict and guaranteed respect to privacy and confidentiality for mobile phone subscribers.

In this line, joint working groups including public institutions (statistical offices, mainly), MNOs, and European and national Data Protection Agencies are recommended. Both interrelated strategic and technical issues are to be taken into account.

- Mobile network data represent an extremely sensitive data source both for private and public statistical purposes. Indeed, access to this data source has implications beyond the production of official statistics, since the perception of privacy and confidentiality loss may easily arise especially under a lack of transparency. Regarding the production of official statistics, this demands extra attention by statistical offices, but also for the public sector in general. In our view, this kind of sensitive data needs a close collaboration within the public sector with a leading and coordinating role by NSIs, ideally imbricated in the respective European and national data strategies.

- As a natural derivative of the preceding item, a detailed communication policy is needed. On the one hand, this communication policy must transmit in an open and detailed way what data are accessed, how data are accessed and processed and who access and process these data. On the other hand, this communication policy must be created and coordinated both at national and European levels. Mobile network data are generally produced by multinational companies with standardised technologies and similar social contexts in the ESS. This clearly favours the European dimension of this data source. We believe that an extra effort in the ESS to create and disseminate a shared communication policy abounds in the benefits of all actors and of society, in general. Some specific elements of this communication policy are suggested below.

- As stated in the preceding sections, we observe the natural crucial role of Data Protection Agencies regarding both the access and processing of mobile network data. This is related to the privacy and confidentiality issues mentioned above as well as the legal implications regarding the protection of personal data. We recommend two courses of action. On the one hand, a research track on statistical disclosure control and privacy-preserving production processes should be prioritised investigating the trade-off between risk and utility in scientific terms taking into account results in the international community (see e.g. Tu *et al.*, 2018; Zang and Bolot, 2011). On the other hand, a specific part of the aforementioned communication policy should be dedicated to openly disseminate and communicate the application of these statistical disclosure control methods, with special attention to DPA officials and citizens in general.

- In this project we have proposed an end-to-end statistical process from raw telecommunication data to final statistical estimates (present population and origin-destination

matrices in our example). Given the volume and velocity of generation of mobile network data, we view in-situ processing (at MNOs' information systems) as the best and more realistic choice. Details about what level of aggregation is appropriate for transmission to NSIs for further processing and how to combine data from multiple MNOs (e.g. with privacy-preserving technologies) are still a matter of research. Details about the end-to-end statistical process and its implementation in production must be further worked together jointly between NSIs and MNOs.

- The statistical methodology used in production should be an element of the communication policy providing details about the nature of input and output data and statistical methods used at each production step.

- In close relationship with the statistical methodology, the use of software tools and the dissemination of open source code should be promoted in line with the communication policy. A balance between the open dissemination of production tools as a high standard for Official Statistics and intellectual property rights and industrial secrecy must be found. Trust from society should be a goal pursued by a widely open exercise of transparency.

- An unsolved issue regarding the access and processing derived from the preceding items is found in the associated costs. This data source and its processing requires specialized staff, software, and hardware. Administrative instruments according to legal regulations (public contracts, public tenders, . . .) are to be used to identify and agree how these costs are to be dealt with. The issue about the costs is delicate, since data for the production of official statistics must be provided for free. This golden principle must be respected also for new digital data.

  In this line, can data collection operations with traditional survey data be used as an inspiring example? Dedicated companies are contracted through public tenders under openly disseminated technical criteria. Strict guidelines and training are provided both for interview administration, data collection and data editing during collection, i.e. NSIs provide a (sub)process design which is executed by external agents. Further work is needed in this direction for mobile network data.

- A critical issue in setting up partnerships between NSIs and MNOs and reaching a collaboration scenario is the delimitation of both public and private product scopes. Consider, for instance, the dissemination of present population counts. If NSIs produce these present population counts as an open public statistical output at the geographic level of, say, postal codes every hour over the whole national territory, the space for private products is reduced. There exist statistical products of null public interest (e.g. population density at specific days around specific retail stores) which can constitute a flourishing activity in a digital economy. European and National Statistical Plans should consider, potentially even through the respective sectorial statistical regulations, detailed accounts of the public statistical products together with their breakdowns.

- In terms of internal strategy within the ESS and National Statistical Systems, we detect a potential risk and obstacle to reach a collaboration scenario with MNOs in the proliferation of isolated business cases, not only with this data source but also with other privately held data sources. This is intimately connected to the everlasting debate about production in silos in the NSIs. Traditional production was organised (still is in some NSIs) independently around each statistical product or statistical domain. Process design and implementation, methodology, software, etc. were not shared and scaled. International production standards [REF] were proposed and implemented to overcome this situation

within each statistical office and even among different statistical offices. With the advent of digital data, we run the risk to reproduce a similar situation if for each statistical domain or each statistical product concrete local agreements for access to specific data sources are reached without a harmonized plan to bring them in a process-oriented production framework.

As a final thought-provoking reflection, let us comment on a well-known and widely used financial indicator called EBITDA, i.e. Earnings Before Interest, Taxes, Depreciation, and Amortization. This is a key indicator in business science to assess the value of a company. Notice how the concept of tax is fully assimilated. A tax is a fraction of the results of the economic activity transferred to the State for the Public Good under highly specific legal regulations. Shouldn't data receive the same consideration?

Part II

Collaboration scenarios

Empirical assessment of principles of access and scenarios with MNOs[1]

Mobile network user localisation data constitutes a valuable data set from several points of view. For NSIs, if access to these data is possible, it constitutes a new source of information to produce statistics of public interest. For mobile network operators (MNO), there is a possibility to develop, on a commercial basis, a custom-made offer of statistics of the flow of people in specific locations.

These data sets demonstrate three distinctive characteristics for NSI:

(i) [market definition tension] The competing use of data to produce statistics of public interest and/or commercial taylor-made statistics creates, prima facie, a tension between NSI and MNOs.

(ii) [more than secondary data] The raw data are issued from an industrial process, the operation of the mobile network in this instance, and thus needs to be refined in order to be usable by official statisticians. This is a significant change compared to the traditional model under which data collected form companies is usually easily derived from the accounting system or from an administrative regulation to which the company is already required to comply with.

(iii) [regulation] The information enshrined in the raw data, i.e. the location of individuals, is highly sensible from a privacy point of view and as such is subject to a double legal standard: compliance with the GDPR[2] and compliance with the e-privacy[3] directive.

This document explores the issues that need to be addressed for NSIs to access such data (chapter 6), and describes three possible scenarios to implement such access (chapter 7).
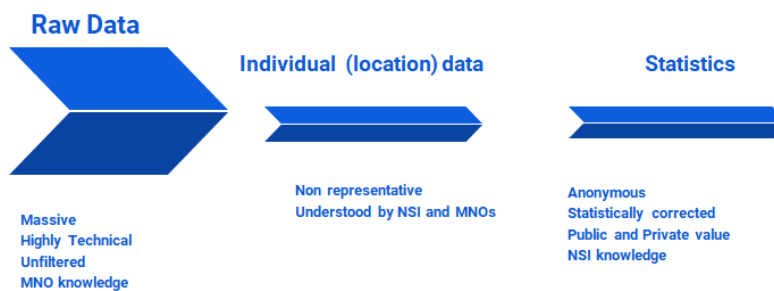
---

[2]Regulation (EU) 2016/679, requiring in particular data minimisation and information about the purposes of the treatments.

[3]Directive 2002/58/EC states that metadata must be anonymized or deleted if users did not give consent. But the future ePrivacy Regulation currently under discussion could relax partly these legal provisions.

Accessing mobile network user location data: framing the issues

Over the recent years, MNOs have been actively promoting the use of their dataset to answer many questions of public or private interest involving the counting of individuals in specific locations or analysing their mobility. The business units producing statistical outputs have mainly four types of clients: the transportation sector (e.g. bus companies), the retail sector (e.g. shopping malls, amusement parks), the local authorities (e.g. municipalities) and the local economic actors (e.g. tourism development agencies). For instance, the municipality of Megève (France) has been working with FluxVision for several years to determine the number of arrivals and departures of visitors, of excursionists, in transit, seasonal workers, tourists...[1]

The current Covid-19 crisis has brought new possibilities to use this data to assist sanitary authorities. The drastic decrease of tourism activities, the significant reduction of higher education services or the imposition of stay-at-home measures have triggered significant movements of population across all territories. Potentially, those movements may call for an adjustment of the capacity of the provision of sanitary services (emergency services, provision of vaccine, ...). Measuring, on an aggregate basis, the decrease in the mobility of individuals in their vicinity can provide valuable information to modelize the impact of specific measures aiming at reducing the spread of the virus.

Three issues seem particularly salliants and need to be addressed in order to consider that NSI may produce new public statistics with mobile network generated data.

6.1. Documenting the methodologies of MNOs to produce: an issue of public policy and a requirement for a sustainable commercial success

The development of statistics deriving from MNOs' operation is still nascent. It is generally characterized by the limited information available on the treatment being implemented by MNOs to deliver intermediate data for statistician or customized aggregate data for end-users[2]. To our knowledge, MNOs usually provide some documents on methodology but: 1/ MNOs' methods of data adjustment to the total population are not disclosed (each MNO exploits its own data but aims to give figures that are representative of the total population.) 2/ for some MNOs, the developed methodology is considered both a legal and an algorithmic innovation

---

[1]See 2018 activity report.
[2]See section 7.2.

and should be therefore subjected to business secrets[3].

Thus, it is extremely difficult to assess both from a theoretical or empirical point of view the relevance and accuracy of the information produced and to detect any bias resulting from the collection process or from the initial treatment of the data. This drastically restricts the ability to use these statistics in any public or private decision process. Decision makers have no way to assess whether they are crossing the limits of validity of the statistics.

From a public policy point of view, there is a public interest to request private statistics providers to release comprehensive metadata to ensure that the statistics are used appropriately. Likewise pollsters are required to disclose key information on any poll produced (sampling method, non-response rate, etc.). From a commercial point of view, it is difficult to imagine how the market of mobile statistics can develop without clarifying the value of the statistics sold in a transparent and credible fashion.

The tension between the need to protect innovation and to establish the value of the product sold is common. It is solved through a mix of limited disclosure and the establishment of a sound governance scheme allowing end-users to put faith in the information released. Typically a trusted third-party, either a public body and/or a private auditor, with an extended access to non-public information, can enhance the credibility of the disclosed information.

NSI are clearly well positioned to help solve this problem. To produce public statistics, they share, with subsequent end users, the need to understand how the raw data were apured, how the statistics were adjusted from any bias arising from the collection process or from the structure of the client base of the MNO.

Data preprocessing at the micro level is key for the final quality of the statistics. Some form of access to this data preprocessing step for the NSI is needed in cooperation with the MNO expertise. It can create a shared value by improving the relevance and accuracy of both the commercial and public statistics produced by the MNO and by the NSI.

To the extent that MNOs are open to collaborate with them, the NSI can act as a trusted third party to review the relevance of the methodologies, the scope of disclosure and thus enhance the credibility of the disclosed information. NSI commonly access and protect from disclosure commercially sensitive information. In this instance, NSI could review the whole methodology implemented by the MNO to produce the individual locations and the treatment implemented to adjust the statistics produced. This review, which would qualify as an audit of the individual location data produced by the MNO, is needed to ensure an appropriate subsequent statistical treatment by the NSI. It would also allow the NSI to act as a facilitator, assisting MNOs in identifying which information should be disclosed for end-users.

Just like auditors do not certify the accuracy and relevance of financial statements, NSIs should only provide comfort on the relevance of the treatment implemented and the MNOs should bear the full liability arising from the production of commercial statistics.

---

[3]"The promising mobile-phone records are hindered by issues concerning availability and consistency. [...] Although various developments are underway to overcome these limitations (e.g., algorithms to correct data biases, switching from Call Detail Records to Signaling data, and trusted smart statistics frameworks to harmonize access to data from different operators), substantial technical and organizational progress is still required for a more systematic use of these data" in Batista e Silva et al. (2020).

6.2. Establishing the legality of the data collection and processing

The information enshrined in the raw data, mobile users whereabouts, is highly sensitive from a privacy point of view and is thus subject to strict legal standards under the GDPR framework. In addition, electronic communication service providers are subject to strict requirements in terms of privacy protection, which preexist the GDPR regulation, through the E-privacy directive. A revision of this directive has been under discussion with the EU co-legislators for many years and is now urgently needed.

The compliance with GDPR to produce aggregate statistics, both for NSIs or commercial ventures do not raise unsolvable questions. The nature of the data calls for specific measures to assess and mitigate the privacy risks.

Privacy protection risk minimization principle entails that the minimal set of data should be transmitted to the NSI for its statistical processing. NSIs are trusted with highly sensitive data in other domains and are complying with GDPR and public statistic laws. Controls over computation at the (individual) micro-level may be insured by (i) transmission of pseudonymized micro-aggregates (ii) access on MNO premises to pseudonymized micro-aggregates by NSI statisticians in charge of calibration, aggregation (iii) specifying to MNOs the computations which would lead to anonymous aggregates for them to implement.

Compliance with the E-privacy directive is more complex as the production of aggregate statistics isn't clearly stated as an authorised use of raw data. The draft being discussed in late 2020 of the new directive includes a provision explicitly stating that producing statistics, in strict compliance with GDPR, is a legitimate use of networks raw data. Such clarification is clearly needed to ensure a sustainable use of MNO raw data for statistical purposes both for public and commercial purposes.

In any case, as part of the compliance with the requirement to protect the privacy of mobile customers, MNOs should disclose the core element of their methodology and the measures being implemented to protect the privacy of their customers.

6.3. Understanding the economy of the data access

The production of individual location data from raw data extracted from mobile network operation has a cost: conceptual investment to design the data extraction and processing systems, IT implementation cost, qualification cost.

If the MNO decides to monetize its dataset, the price paid by commercial end users covers this cost augmented by a commercial margin. This margin reflects a combination of the intrinsic value of the data (a by-product of mobile network operation) and the value added by the MNO. In such a situation, the request from the NSI to access information of general interest should follow general principles: 1/ free access as a general principle ; 2/ monetary compensation in rare exceptions when the incremental costs deriving from addressing the NSI request constitutes a disproportionate cost for the venture. Hence, when the MNO has developed a commercial activity, it is unlikely that the NSI should provide monetary compensation as the economy of the commercial activity bears the costs of producing the raw data.

This analysis has a corollary: the statistics produced and disseminated by the NSI should marginally infringe the commercial value of the data set. NSI statistics are public, can be reused

by anybody and thus have lost any commercial value. Tensions between the MNO and the NSI may arise on this specific issue and should be addressed upfront. NSI should be ready to make credible commitments to strictly limit the scope of the statistics produced. In collaborative scenarios between MNOs and NSI, such commitments should be legally binding for NSIs.

If the MNO hasn't developed any commercial activity to monetize this data set, then the cost of extracting and refining the raw data is entirely incremental and attributable to the NSI request and should be further analysed to ascertain whether it constitutes a disproportionate charge for the company warranting a monetary compensation.

In any case, in accordance with existing legal frameworks, NSIs should not consider providing monetary compensation based on the 'intrinsic' value of the data. NSIs should recognize that the data accessed is being developed for commercial purposes and be careful not to impact the dynamic of this nascent market by strictly limiting the scope of the public statistics produced and disseminated. NSIs should also demonstrate a strict neutrality with all market players, i.e. be open to work with any interested MNO and follow strict procedure to prevent any transfer of the information received from one market player to another.

Analysis of different scenarios according to the identified challenges

7.1. Scenario 1: Mandatory access

A first solution to ensure NSIs' access to MNOs' data is to make the transmission of MNOs' data mandatory by law. This solution has already been implemented in some countries for other accesses to privately held data (see Box 1). It should be made clear that these legal provisions would be in addition to the necessary legal provisions on NSIs' access in the E-Privacy Regulation. The mandatory access falls within the scope of the Statistical Law whereas the authorized purposes to access are subject to regulation of the electronic communication sector. Almost all scenarios in this document require the introduction of both commercial and public statistical purposes in the list of purposes allowing the processing of metadata in the E-Privacy regulation. However only this scenario of mandatory access requires another legislative provision.

Box 1: An example of mandatory access to scanner data in order to compute the CPI index in France

In France, thanks to the Article 3 bis of the national Statistical Law[1], the Minister of Economy can decide that some privately held data have to be transmitted electronically to the public statistical service, under some conditions (public interest and feasibility proof notably). For the computation of the Consumer Price Index (already an official statistic), a decree was published in April 2017, which requires all retailers to transmit their scanner data to INSEE. Since the beginning of 2020, the integration of these scanner data in the CPI process has allowed INSEE to produce new statistical products and improve the quality of the standard CPI. They also proved to be particularly useful during the Covid-19 crisis because of the cessation of on-site price surveys.

This solution using the legal constraint requires mainly two conditions. First, the political support should be strong. The public interest of the final statistics should be clear to all stakeholders and supersedes the tension around using individual mobile network metadata for statistical purposes. Second, the requested data should correspond to unambiguous measures: e.g. sample of events relative to each mobile phone user or counts by antenna. Because the law must apply uniformly to all MNOs, a clear definition of the scope of the data is required,

---

[1] https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000888573/2020-09-17.

without any reference to the specificities of the network, which could change over time or be different between mobile network operators. In theory, this scenario enables the consolidation of all MNOs datasets.

We now distinguish between a mandatory access to individual location data and a mandatory access to aggregated data and we identify the strengths and weaknesses of both scenarios.

7.1.1. Mandatory access to individual location data

We refer, for instance, to sample of events relative to each mobile phone user.

These individual location data represent a very large amount of data: according to the estimates of different MNOs, the total number of events per day is assessed to several million. Moreover, this number is likely to change (and more likely to increase) as technologies improve and user behaviour changes. But this very large volume of data is not in itself an obstacle for NSIs whose job is already to handle big data. However, the transmission of such a volume of data will generate significant costs to adapt the IT infrastructure that the NSIs must be prepared to bear.

The systematic transmission of all individual data also requires addressing the crucial issues of data security and privacy. NSIs have already faced with the issues of personal data and more recently, with private data issues. NSIs' infrastructures and the practices of public statisticians are already in compliance with the rules ensuring data security and privacy. Nevertheless, the request for individual data represents an additional risk to the NSIs compared to the transmission of aggregated data only.

In the end, the main obstacle would rather be that of the proportionality of the request. Some preparation of data would be necessary from MNOs, which do not have a commercial activity to monetize these data. Even if the requested data are unambiguous and may appear as basic, the treatments, the storage and the production of metadata needed to inform the production process can represent a noticeable workload and induce an incremental cost for the MNO infrastructures. Moreover, because of the complexity and the need to use scarce data science skills, the MNO cooperation is likely to be limited.

On the other hand, the access to individual data allows NSIs to construct and to communicate its own methodology concerning the construction of statistical indicators. In particular, this would make it possible to better meet the needs of the users of official statistics and to improve the quality of final indicators by combining sources at the finest level (see Box 2). These criteria of independence, quality, transparency and adequacy with the needs are part of the recommendations on Good Statistical Practice adopted by the NSIs.

Box 2: Combination of sources with individual location data

The analysis of mobility could be enriched with socio-demographic data at the individual level. For example, mobility surveys today make it possible to report on mobility by age, gender or even socio-professional category. These elements are essential to better design the future public policies related to transport or mobility.

To sum-up, this first scenario would make it possible to construct relevant statistical indicators based on MNOs data. However, it requires two legislative changes. In terms of our evaluation criteria, this scenario is very much in line with the official statistics principles but may seem too ambitious, at least in the short term from an implementation point of view.

7.1.2.   Mandatory access to aggregated data

We refer e.g. to counts by antenna.

The request to access only anonymous aggregated data allows NSIs to avoid the sensitive debate of access to individual data. In this case, privacy protection would remain the domain of the MNO only. This also avoids the inhouse infrastructure costs of hosting very large individual data sets for NSI.

But this way of operating data (aggregation on MNOs premises) will probably be an obstacle to obtaining the most relevant indicators. In this situation, the combination of sources is indeed very limited. Consequently, the outputs will likely not be as relevant as expected.

In addition, because of the absence of control on quality and process before the data transmission, the processing process will be far from meeting the criteria laid down by official statistics. Indeed, the necessary involvement of MNOs should go beyond access: consistency of the aggregated data checks, support and dialogue with the NSIs but these cooperative behaviours are unlikely in a situation of constraint.

7.2.   Scenario 2: Access by commercial contracts: pay MNOs to implement statistics corresponding to NSIs' needs

Some public actors resort to MNO's statistical products through commercial contracts usually for specific and local case studies. As of today, they are examples of MNOs offering similar modalities to NSIs for acquiring data. This scenario based on a commercial contract seems in the short term the more realistic to ensure access to NSIs as it requires no effort from the MNO side, rather the converse. However, it seems also the most unsatisfactory given the most basic statistical public mandate. This scenario should rather be considered as a status-quo driven by MNO private interest and data monopoly which can not foster a long-term statistical production.

7.2.1.   Existing statistical offers

In principle, existing statistical offers have no reason to match NSI's needs: on top of the specificity of NSI's demand (e.g. present population vs places' frequentation), calibration may require matching with NSI's own data sets to ensure representativeness. In addition, MNO's statistical products are usually not intended to be handed over to statisticians but rather to the final end users. In commercial offers, emphasis is on the final output (e.g. through dashboard, visualisations...) and entry points are more often product managers than data scientists. In turn, dialogue with NSI's should favor methodology, quality, integration of statistical processes as NSI are accountable for the final quality, meta data and interpretation of the statistics they disseminate.

7.2.2.   Specifying the needs of the NSI, collect and pay for aggregates

To answer limitations of existing statistical offers, NSIs could specify their needs and develop a contractual relationship with MNOs to implement an adapted solution. As for 2.1, NSI data acquisition process is then dependent on a commercial relationship, a setup which changes the paradigm under which NSIs currently operate as defined by existing statistical laws, i.e. NSIs should not provide any monetary compensation when the data collected is needed to produce statistics of public interest and provided that the cost incurred by the companies remains proportionate. If adopted on a larger scale, such a commercial model would set a dangerous precedent for the sustainability of public statistics in a future datafield society where increasingly more data are in private hands. An NSI can not pay for data market value nor to cover MNOs investment to monetize their datasets.

Yet, whatever the scenario, some data operations will necessarily be executed by the MNOs. This scenario considers the case where some form of monetary compensation for data treatment cost executed on behalf of the NSI could be agreed upon based on a contract. A sharp distinction will be needed between compensating the incremental cost of the data managing work born by the MNO and paying as regular customers. Arguably, evaluating the incremental cost compensation beyond a proportionate demand falling within statistical laws will be difficult, and information will be asymmetric. In addition, it ignores the benefit an MNO can get at involving an NSI into its statistical process, mainly but not only as a reputation landmark. Most likely, this scenario will fall short of NSI's expectations.

Such a model is better described as a two phase production rather than a real combination of NSI and MNO data processes. In this model, the NSI is not in a position to guarantee the quality of the first phase. The NSI may lose the trust of its end users by relying on private processes with limited control and by putting himself in a position of dependence from a few private actors. If the NSI restricts its role to specification of the statistical process, without actually implementing it and without safeguards controls, it changes the role of the NSI as a mere dissemination role which should be at least questioned. If the MNO private statistical business should be legitimately recognised, the mandate of the NSI should be also recognised as well in a balanced partnership with appropriate safeguards.

7.3.   Scenario 3: Access through a win-win partnership

Society as a whole will benefit from a better and more relevant use of mobile phone data statistics. These data are underused today despite their potential because business models have difficulties to emerge and reach a limited number of actors. MNOs have already shown their consideration for public interest notably during the COVID crisis, and are used to collaborate with public authorities given the public-interest importance of their economic activity. To preserve the legitimate business interest of the MNOs while producing public interest statistics with the appropriate safeguards (trust of the end users, privacy preservation, independence and objectivity), a novel form of public-private collaboration should be considered. To ensure that both parties can benefit from the partnership, NSIs should accept, through a covenant of a partnership agreement, to reduce the scope of their dissemination to a general and low-frequency structural information, useful in priority for the clearest public-good use cases and for stakeholders not regularly popping within MNOs client portfolios. It should be noted that these stakeholders exist. For instance in France, local sanitary and environmental agencies have expressed the need for a structural seasonal present population statistics for sanitary, risk surveillance or even waste management. High-stake public needs which are to our knowledge not addressed today

by MNOs statistical products are low-frequency but comparable over long period statistics. Needless to say, high-stake implications require an accountable and transparent producer - a role which should not be undertaken by MNOs in democracies. This effort in dissemination scope definition intends to preserve the interest of the partner in terms of business outcomes. This should be recognized by the NSI as essential for a sustainable and efficient model of co-existence.

The NSIs have several things to offer, as a trusted and qualified third party. The qualification of the relevance of MNO's statistical treatment or the participation to their elaboration can significantly increase the trust of the MNOs clients and resolve the short-comings of closed methodologies. NSIs have both expertise and access to high-quality calibration data. Calibration data are valuable and often accessed for free in their open-access version by MNOs. Calibration for representativity, correction of selectivity-bias remains the domain of competency of NSIs, which could be shared with MNO in an integrated join-production process. Note that representativeness is a key element for guiding fair public policy.

Once the dissemination scope has been agreed, dissemination of trusted structural statistics could foster an increase in demand for short-term, granular or tailored case studies. As the competition risk would have been designed to be minimal, MNOs would most likely benefit from the large-audience dissemination of official statistics built on their data. By making clear the public good effort, MNOs would probably gain in reputation.

In such a cooperation, the MNO should recognize the legitimate mandate and skills of the NSI, its commitment to transparency, objectivity and independence. If the MNO is committed to producing reliable statistics as their feedback tends to indicate, we argue it should not be hard to reach an agreement on the necessary information to deliver to the end-user to understand and objectively assess the strengths and weaknesses of the data.

Discussions on data management and workload should be done under a common understanding agreement:

- Each party should bear its own cost and recognize that the purpose of the partnership is a mutual benefit for each party and not the provision of a commercial service to the NSI. For instance, if the adjustment of the mobile phone data required the calculation of weights by the NSI, the NSI would bear the cost of producing this weighting.

- The NSIs recognize that MNOs are the data holders and pursue a legitimate economic goal by producing private-interest statistics, whose existence are key to sustain the model, and that the incremental cost for private actors to address NSI's request should be minimized and proportioned.

- The MNOs recognize the legitimate mandate of NSIs to produce public interest statistics and to access the necessary and proportional data to fulfill its mandate.

We believe such a partnership has the potential to benefit all involved parties and society as a whole, but we foresee that without a strong political support and legal enablers, discussions between parties may be endless.

8

Conclusions

To conclude, two points seem crucial for the future of the cooperation between NSIs and MNOs.

First, the revision of the eprivacy regulation must allow mobile phone data to be used for commercial and public statistical purposes. In the absence of this legal change, NSIs would be constrained to access only aggregated data, without any chance to understand, to qualify and/or to develop the methodology. This seems unsuitable in the long term according to the arguments developed above.

Second, in this paper we have identified two main problems of confidence in the use of MNO data. There is a lack of confidence in the quality of data that prevents public decision-makers from taking full advantage of their potential and there is a lack of confidence from the MNOs' clients in the strict respect of privacy in treatments. In both cases, the introduction of official statistics as a trusted third party would be able to solve the problem to a large extent.

This position could allow NSIs to look further in the methodology than the usual clients. Based on the model of the audit work, NSIs could ensure that adequate metadata would be available without seeking to standardize or harmonize but with the aim of qualifying the outputs. This quality signal provided by official statistics, resulting in a significant increase in the market could be valued by MNOs. Provided that NSI can limit the scope of their statistics produced and published and do not try to address all the need of public authorities , the corresponding gain would be much greater than the potential loss of the market caused by the publication of some structural indicators by NSIs.

In the digital age, understanding of the data process and trust in resulting statistics are necessary conditions for making informed decisions. In a situation of quality enhancement by NSIs, the use of statistics based on MNO data by the public sphere will indeed certainly be more intensive. But in case the methodology remains opaque, even for NSIs, as it is the case today, this rather supports the ending of the use of mobile phone data by public decision-makers.

Today, public authorities use mobile phone data without being able to access the methodology and therefore without being able to qualify their validity. This situation can only be transitory. If conditions were to remain the same, it seems unlikely that the market for the use of MNOs data would develop, to the detriment of both their producers and users.

Appendix A

Template for the Guidelines

In the next page we include the template adopted by each partner in the current WPI to gather feedback from the corresponding national MNOs. This is a template further adapted in each case to the concrete national situation and the concrete collaboration and information with respect to each MNO.

Although appearing as a set of questions, the interaction is thought to be as much personalised as possible in order for MNOs to express their viewpoint on these issues. The questions are intended to serve to statistical officers as guidelines to collect MNOs' points of view. The ultimate goal is to identify the main obstacles for MNOs to share and provide data access for the production of official statistics.

PROPOSED STRUCTURE FOR A DIALOGUE WITH [MNO NAME]

WPI on Mobile Network Data -- ESSnet on Big Data II

[NSI NAME]

[Date]

This document proposes a draft structure for a dialogue with MNOs to collect and gather their feedback on different aspects of the access and use of mobile network operator (MNO) data for the production of official statistics. There already exist different initiatives in the European Statistical System (ESS) and the European Union, in general, regarding the access and use of this data (and Big Data in general) such as preceding works by the ESSnet on Big Data I, the Group on the Use of Privately Held Data, or the expert group on B2G data sharing. Other institutions (such as United Nations) have also produced some material regarding the access to digital data sources in private hands.

The track 1 on access to MNO data of Work Package I on mobile network data in the ESSnet on Big Data II seeks to collect and to gather as much valuable information and feedback directly from collaborating MNOs as possible. The aim is to clearly identify the critical issues *according to their view* so that more work can be planned and carried out by National Statistics Institutes (NSIs), Eurostat, and the ESS in general.

Information about the access to mobile network data clearly has many facets thus potentially involving different departments, units, and profiles in an MNO. The ideal target for this dialogue will be those officers within each company both having a knowledge of both the issues mentioned below and/or having a (partial) responsibility on the decision to share data and work together with NSIs.

This information will be the core of one of the deliverables of the project, released in due date and form through the official dissemination channel of the project. Anonymity regarding the input provided by each company will be respected, especially if explicitly requested by MNOs. The priority is to set up an honest, direct, and fluent dialogue with MNOs so that their standpoints are clearly identified and acknowledged by NSIs, Eurostat, and the ESS in general.

We propose a structure of issues to be dealt with them. These issues aim at potential obstacles to arrive at a statistical production system of official statistics including MNO data in a sustainable way in time. We want to avoid rigid questionnaires, closed answers, and pre-fixed questions, since every situation with every MNO and every country is different. However, we provide a list of items which every NSI can (and probably should) reformulate according to each different situation. The main objective is to collect this information, the wording is not essential, but just the means to achieve this.

This specific instance of the guidelines has been adapted for [MNO Name] based on preceding collaborations. Should any issue that the company considers relevant be missing, please feel free to include it at your best convenience.

PART 1. WHAT POSSIBLE PARTNERSHIPS CAN BE ESTABLISHED BETWEEN NSI [use your case] AND MNO [use your case]?

This part focuses on identifying those common points of interest in the production of statistics between MNOs and NSIs. We need to identify those elements in NSIs which MNOs perceive as attractive enough as to build a collaboration around them.

In our understanding of an increasing interest in the possibility of a "win-win partnership" between official statistics and mobile network operators:

Q1. How do you see the contribution to a collaboration with official statistics for the operator?

Q2. Would you be willing to enter into a research partnership with NSI [use your case]? If not, why? If conditionally, what would these conditions be?

Q3. Would you consider improving the quality of statistics produced by MNO [use your case] using additional data from NSI [use your case] an asset for the operator?

Q4. If a common methodology is developed for the production of statistics of public interest, would open source software be an impediment? What obstacles do you identify?

PART 2. INSIGHTS FROM MOBILE NETWORK DATA

This part focuses on identifying those statistical insights more valuable for MNOs. We need to identify what kind of statistical products are of primary interest for them to analyse potential alignments. To design a win-win parternship, the statistical product produced by the NSO should be clearly identified, in the interest of end users and carefully with respect of MNOs business interest.

Q1. Mobile phone data is used to produce statistics of private and public interest. Could you identify what, in your opinion, is the main added value of mobile network data compared to other available sources of information (surveys, administrative data, population census, etc.)?

Q2. What are the different types of target customers of your statistical product portfolio? For these customers, can you outline the statistics provided/demanded? Do you know their use (operational management? Analysis? Risk measurement?)?

PART 3. BUSINESS INTERESTS

We are now interested in the business interests pursued with the production of statistics by the operator.

Q1. Does MNO [use case] develop an economic activity around the production and commercialization of statistical products based on mobile network data?

A. In the positive case:
> **Q2a.** Could you briefly describe the history of your business and commercial activity around the production of statistics from mobile network data? [Suggested items: Date of creation, international development, turnover, growth prospects, risks, periods of weakness, etc.]
> **Q3a.** Would you be ready to initiate consultations on the statistics that might be produced jointly with NSI [use your case] in order to find a balance between your commercial interests and the production of official statistics?

B. In the negative case:

Q2b. Could you briefly describe the reasons leading to this decision?

PART 4. LEGAL ISSUES

Q1. According to the e-privacy directive [and its counterpart in the national law, if applicable], the use of the postal code (or similar) and electronic communications is restricted regarding the usage of mobile network data. Has this regulation limited or is this regulation limiting your activities of the operator in the field of statistical production?

Q2. Regulation No. 2016/679, known as the general data protection regulation (GDPR), was released in 2016 and strictly regulates the use of personal data, conditioning it in particular on the consent of citizens. Does this regulation constitute a burden or even an impediment on the economic activity of the business unit monetising mobile network data?

Q3. Are the anonymization rules enacted by the national Data Protection Authority [DPA; use the concrete name in each national case] a burden for the operator? Are anonymization processes expensive? Has the operator already been audited by the national DPA?

Bibliography

Batista e Silva, F., S. Freire, M. Schiavina, K. Rosina, M.A. Marín-Herrera, L. Ziemba, M. Craglia, E. Koomen, and C. Lavalle (2020). Uncovering temporal changes in Europe's population density patterns using a data fusion approach. Nature Communications 11, 4631 (2020) doi:10.1038/s41467-020-18344-5.

de Montjoye, Y.A., C.A. Hidalgo, M. Verleysen, and V.D. Blondel (2013). Unique in the crowd: The privacy bounds of human mobility. *Nature Scientific Reports 3*, 1376–1381.

DGINS (2018). Bucharest Memorandum. `http://www.dgins2018.ro/bucharest-memorandum/`.

Galiana, L., M. Suarez-Castillo, F. Sémércube, E. Coudin, and M.-P. de Bellefon (2020). Retour partiel des mouvements de population avec le déconfinement. Technical Report 54, INSEE. INSEE Analyses.

Gregoir, S. and F. Dupont (2016, March). La réutilisation par le système statistique public des informations des entreprises. Rapport du groupe de travail INSEE-CNIS.

INE (2019a). Anuncio de formalización de contratos de Instituto Nacional de Estadística. Boletín Oficial del Estado, de 21 de enero de 2019. Expediente 2019N0060001. `https://www.boe.es/boe/dias/2020/01/21/pdfs/BOE-B-2020-2497.pdf`.

INE (2019c). Anuncio de formalización de contratos de Instituto Nacional de Estadística. Boletín Oficial del Estado, de 31 de diciembre de 2019. Expediente 2019N0060002. `https://www.boe.es/boe/dias/2019/12/31/pdfs/BOE-B-2019-56547.pdf`.

INE (2019c). Anuncio de formalización de contratos de Instituto Nacional de Estadística. Boletín Oficial del Estado, de 31 de diciembre de 2019. Expediente 2019N0060003. `https://www.boe.es/boe/dias/2019/12/31/pdfs/BOE-B-2019-56548.pdf`.

INE (2020a). Anuncio de formalización de contratos de Instituto Nacional de Estadística. Boletín Oficial del Estado, de 15 de junio de 2020. Expediente 2020N0060002. `https://www.boe.es/boe/dias/2020/06/15/pdfs/BOE-B-2020-16961.pdf`.

INE (2020b). Anuncio de formalización de contratos de Instituto Nacional de Estadística. Boletín Oficial del Estado, de 9 de octubre de 2020. Expediente 2020N0060004. `https://www.boe.es/boe/dias/2020/10/09/pdfs/BOE-B-2020-34257.pdf`.

INE (2020c). Anuncio de formalización de contratos de Instituto Nacional de Estadística. Boletín Oficial del Estado, de 20 de noviembre de 2020. Expediente 2020N0056003. `https://www.boe.es/boe/dias/2020/11/20/pdfs/BOE-B-2020-42528.pdf`.

INE (2020d). Anuncio de formalización de contratos de Instituto Nacional de Estadística. Boletín Oficial del Estado, de 10 de diciembre de 2020. Expediente 2020N0056002. `https://www.boe.es/boe/dias/2020/12/10/pdfs/BOE-B-2020-46213.pdf`.

INSEE (2020a, June). By the end of May, morning commutes had only reached 60% of their usual level. https://www.insee.fr/en/statistiques/4620435?sommaire=4473307. INSEE Economic Outlook.

INSEE (2020b, May). Population présente sur le territoire avant et après le début du confinement: résultats consolidés. https://www.insee.fr/fr/information/4493611. Press release of May 18.

INSEE (2020c, April). Population présente sur le territoire avant et après le début du confinement: résultats provisoires. https://www.insee.fr/fr/information/4477356. Press release of April 8.

Verhulst, S.G., A.J. Zahuranec, A. Young, and M. Winowatan (2020). Wanted: Data Stewards: (Re-)defining the roles and responsibilities of Data Stewards for an age of data collaboration. The GovLab, March 2020. `http://www.thegovlab.org/static/files/publications/wanted-data-stewards.pdf`.

WP5 of ESSnet on Big Data I (2016). Deliverable 5.1. `https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/6/65/WP5_Deliverable_1.1.pdf`.

WP5 of ESSnet on Big Data I (2017). Guidelines for the access to mobile phone data within the ESS. Deliverable 5.2. `https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/6/65/WP5.Deliverable1.2.pdf`.

Tu, Z., F. Xu, Y. Li, P. Zhang, and D. Jin (2018). A new privacy breach: user trajectory recovery from aggregated mobility data. *IEEE/ACM Trans. on Networking 26(3)*, 1446–1459.

Zang, H. and J. Bolot (2011). Anonymization of location data does not work: A large-scale measurement study. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, MobiCom '11, pages 145–156. Las Vegas, Nevada, USA.