

**WORKSHOP ON PUBLIC – PRIVATE PARTNERSHIP FOR
MOBILE PHONE DATA FOR USE IN OFFICIAL STATISTICS**

Luxembourg, 22-23 September 2016

MINUTES OF THE MEETING

Participants:

Mobile Network Operators

Orange Labs (France)
OTE (Greece)
POST (Luxembourg)
Proximus (Belgium)
Telefonica (Spain)
Telenor (Norway)
Vodafone (Germany)

National Statistical Offices

Central Statistical Office of Poland
DESTATIS
INSEE
INSSE
STATEC
ONS (UK)
Statistics Austria
Statistics Belgium
Statistics Finland
Statistics Netherlands
Statistics Spain (INE)

International public organizations

DG Connect
DG Digit
Eurostat
Joint Research Center
United Nations

Private sector

Positium

THURSDAY, 22 SEPTEMBER

SESSION I. PRESENTATION: CONTEXT AND ACTORS

1. WELCOME ADDRESS AND PRESENTATION OF THE ESSNET ON BIG DATA

Statistics Spain (INE) opened the workshop by welcoming the participants and expressing their gratitude to Eurostat and STATEC as co-organizers of the event. The workshop was presented in the framework of the ESSnet on Big Data project as a planned activity of the work package 5, devoted to mobile phone data, one of the most promising Big Data source. However, INE stressed the fact that official statistics have to face multiple challenges in the road to investigate the potentialities of these data for the production of European and national statistics. Among them, the construction of a collaborative and strategic partnership with the private sector, and find a space of understanding in which the interests, concerns and needs of official statistics and private companies can coexist.

STATEC, host of the event, also welcomed the attendees to then present the agenda and invite to an active and intense participation in the different sessions.

Finally, Eurostat took the floor to again acknowledge the assistance of all and make a presentation about the mission and values of Eurostat and the European Statistical System (ESS). The Vision 2020, a strategy for the modernization of the ESS, was explained, focusing on the five key areas in which common action is needed in order for European statistics to be “fit for the future”. This can be summarized in (i) meeting users’ requirements, (ii) facing methodological and technological challenges and (iii) innovating its offer in terms of products and services. Data revolution brings opportunities and challenges, and the ESS has to respond to this change of paradigm using all the relevant sources to produce knowledge preserving always our high commitment to quality. To be successful facing all these challenges, the statistical community cannot work alone, we need to build partnerships and to seek strong political support and investment.

2. DIGITALISATION OF HUMAN ACTIVITY AND OFFICIAL STATISTICS.

Eurostat started the agenda by talking about the data revolution in the context of official statistics. The digital footprint that humans left behind can be used to improve statistical production. The Scheveningen Memorandum on Big Data, agreed by the ESS in 2013, called for the submission of an action plan and roadmap on Big Data, adopted a year after, in September 2014. The main aspects of this action plan are the integration of a Big Data strategy for statistics into an overall government strategy, the need for skills, the collaboration at European and global level, the need for methodological developments, quality assessment and IT, the search for partnerships between different stakeholders (government, academics, private sector), communication and the protection of privacy and personal data (legal and ethical issues).

In 2012-2014 Eurostat carried out a feasibility study on mobile positioning data for tourism statistics. From this study an analysis of the strengths and obstacles in the use of big data sources was derived. Among the opportunities a better density in terms of space and time, as well as higher coverage and observation of behavior can be mentioned. On the other hand, problems are encountered in issues as selectivity and bias, lack of

additional information to study socio-economic characteristics, integration of different sources and their sustainability.

A first set of challenges for the future refers to a common and scientific approach of the ESS to the use of big data sources based on cooperation and exchange of best practices. To address this issue, Eurostat is launching a series of pilot projects under the framework of the ESSnet on Big Data. These projects are an important pillar of the big data activities in the ESS in the coming years and should pave the way towards the integration of big data sources into data statistical production.

Secondly, new skills and IT infrastructure are also key to successfully move towards the new sources. To this aim concrete actions are ongoing, for instance the “sandbox” environment for big data experiments hosted by the Irish Central Statistics Office in a cooperation between Eurostat and UNECE, among others. On the other hand, the training strategy is under the umbrella of the European Statistics Training Programme. The courses are focus on sources and tools to assess, use and explore big data.

Political and regulatory framework are also important areas to work on, especially related to the access to data. The discussion is not limited to the entry but should include a long term vision to guarantee a continuity of access.

Other important challenges are data security and privacy concerns or how to find a sustainable business model for big data in official statistics.

Eurostat concluded saying that official statistics community is working on big data projects, trying to establish partnerships, exploring possibilities of combining sources, and finding new uses and applications.

SESSION II. EXAMPLES OF ON-GOING INITIATIVES

1. MINING MOBILE PHONE DATA TO RECOGNIZE URBAN AREAS

INSEE gave a presentation on an on-going joint project with Orange Labs Sense and Eurostat. They are currently working on CDR (call detail record) data sets from 2007 (May-October), approved for national authorities to be used for statistical purposes. The project includes two case studies: (i) urban areas detection and (ii) homes detection. These lead to two use cases, at individual level (done in situ at Orange Labs) to use individual trajectories to identify residence, and at aggregate level (done by INSEE) to classify areas. The objective is to infer area type from mobile phone patterns (residential vs working) and also to compare a new source with a reliable reference (census data 2008).

The processing for urban area prediction was made in three steps:

- 1) To split data in training sample and testing sample. As information is very spatial correlated it has to be first decorrelated.
- 2) To estimate a prediction of antenna’s classification on the training sample and calibrate the algorithm (using several methods and different metrics). Benchmark mobile phone data using sociodemographic variables coming from official sources (census 2008).
- 3) To analyze the results. They show that accuracy is reasonably good, although the prediction can vary among classes.

The conclusions of this pilot study are the following:

- 1) Joint work between NSI and MNO is crucial to develop expertise in the NSI and to evaluate the potential use of this source in official statistics. This partnership is beneficial to the MNO as a way of validating their data for other business purposes by having the quality stamp of official statistics.
- 2) Aggregated mobile phone data are compliant with sociodemographic information used to classify area, and enough to satisfy information needs.
- 3) Further work is needed to replicate the study on more recent data, in order to control what is going on over time and to explore the stability of the results.
- 4) Further analysis is also necessary at a more detailed level.

2. IN-HOUSE (DISTRIBUTED) AND TRUSTED THIRD PARTY (CENTRALISED) APPROACH FOR PROCESSING MOBILE PHONE DATA: PROS, CONS, AND MARKET SHARE CONFLICT POINTS.

Positium's contribution showed the options for processing mobile phone data based in their collaboration with Statistics Estonia and other countries.

Generally, when talking about mobile data processing three phases can be identified: (i) extraction, (ii) processing, and (iii) aggregating results, and they can be tackled from three different options:

- 1) Distributed (in-house) processing. Data are handled by MNOs from the extraction to aggregation and calculation of indicators, while NSIs can combine the results from several data providers. Under this model privacy protection aspects are solved and less skills are needed for NSI. Moreover, MNOs can use the data also for their commercial aims. In general terms, this option is less expensive for NSIs direct costs. On the other side, NSIs does not have access on raw data and no control over the methodology nor the processing that might be different in any case. Also the entailed burden on MNOs requiring compensation was pointed out.
- 2) Centralized processing. Extraction of the data is done by MNOs and raw data (anonymised or semi-anonymised) are transmitted to NSIs who will be responsible for the full processing chain. This guarantees the use of one single methodology, as well as trusted and controlled results. NSIs have a complete control over methodology and algorithms and there are no burden on MNOs further than the costs of extraction. The main disadvantage lays on the privacy protection that legally prevents this option in many countries. Additionally, technical resources and skilled employees (data scientists) are required. Overall, this option is directly more expensive for NSIs.
- 3) Centralized processing trusted in a third party. MNOs extract the data and are transmitted to a trusted third party (government organisation, university, company...). In principle this option would guarantee one methodology and almost controlled results. There would be no burden on MNOs and less need for new skills for NSIs. The highest risks are the unknown methodology, the privacy protection and the higher degree of complexity of the business model as more stakeholders are involved in it.

Technically all the options are possible, thus the best depends on the business model and regulations.

Another element to take into account is the market conflict that could exist between public interests and the private sector. To illustrate this point, Positium show some examples where the market for data could overlap and bring to light the need to agree on the division between markets.

According to Positium the top barriers to big data adoption are the lack of funding, not enough policy-makers engagement, difficulties on identifying key partners and get access to data, as well as to assure new skills, capacity building and technology.

However, in the European Union the biggest obstacle is regulation and privacy protection that could require changes in European and national laws. Still, it has been stated in several countries that if the processing was done in-house by MNOs and the identifiable data were deleted after the results, it would be acceptable.

In any case, to overcome all these obstacles against, it is important to make the actual return of investment (ROI) for the society in long term more visible.

During the discussion many considerations and concerns about the business model arose. NSIs are concerned about issues such as solving the representativity of data, especially in a decentralized model in which there is no control on methodology, and considering that many MNOs do not see the importance of representativity because their customers do not ask for that.

Moreover, MNOs stressed the need to resolve issues such as security, protection of privacy and reputational risk of customers perceiving negatively the transfer of their personal data. They also highlighted the fact that MNOs can be in different situations: some of them can have a business strategy on big data, some can be developing a commercial line and others have not explore yet the use of big data. For MNOs that have carried out a large investment in technology, innovation and human capital to learn how to extract value from big data is very hard to give away or jeopardize the monetization of that strategic investment. Not to mention the strong competition inside the sector in which sharing or generalizing this knowledge, when the level of development between the companies is unequal, can seriously harm the interests of those companies that already have a big data product on the market.

The debate concluded that all these questions have no clear answer at this time, and they can only be resolved by trying to move forward.

3. VALUE FROM MOBILE PHONE DATA: A MUTUALLY BENEFICIAL PARTNERSHIP BETWEEN A NETWORK OPERATOR AND A STATISTICAL OFFICE.

Proximus presented their collaboration experience with Statistics Belgium as a concrete example of a jointly positive partnership.

First, the barriers and problems identify by Proximus were fairly much the same discussed in the last point, mainly privacy and confidentiality concerns. Big data is still in an innovation stage that requires a great effort of investment for the MNOs to be able to develop a commercial product or service based on mobile phone data. This makes them more sensitive to the exposure of knowledge competition. From NSIs side, the constraints focus on the data access, the lack of big data experience and IT infrastructure and the absence of a specific legal framework.

In this context, joining forces can bring important contributions. MNOs can provide data, metadata, infrastructure and technical expertise that could make NSIs progress in the production of faster statistics, with more detail and better validations, coverage and concepts, as well as to reduce response burden and costs.

On the other hand, NSIs can offer to MNOs geocoded statistical datasets, statistical and methodological learning to apply in commercial products, domain and subject expertise, official quality stamp and an opportunity to improve their credibility and corporate social responsibility (CSR).

Proximus, Statistics Belgium and Eurostat tested this approach with a concrete cooperation case aiming to assess the quality of mobile phone data as a source of statistics. From this experience they had learnt the following lessons:

- 1) Mobile phone data is relevant for official statistical production.
- 2) There is a need for close cooperation between public and private sectors to reach a common understanding of what is (and is not) in the data.
- 3) The need for complementary skills and resources.
- 4) Combination of datasets is a necessity, protecting data privacy.
- 5) This kind of partnership gives MNOs external and internal visibility.
- 6) It is positive to start with small and pragmatic projects.
- 7) To get a trusted and solid cooperation is important to focus on mutual benefit, there are many spaces for a win-win.

The value of this pilot study was widely recognised, although some MNOs expressed that this kind experiments are acceptable in a pilot environment but many concerns appear when we think on moving to production.

4. ROUND TABLE: LESSONS FROM THESE EXPERIENCES

Eurostat acted as moderator of the round table composed of the following panellists: Statistics Belgium, Proximus, INSEE, Orange Labs, Statistics Finland and Telenor Norway.

Eurostat opened the round table by bringing up two questions that acted as a common thread for the discussion:

- 1) What are the good lessons (opportunities) and lessons to avoid (risks) in order to build a more solid partnership?
- 2) Concerning data access level, what are the general lessons learned and guidelines to follow?

The main relevant comments shared by the panellists and other participants are listed below:

- 1) An outstanding priority placed on the implementation and development of good and clear use cases to show real usefulness, feasibility and mutual trust was agreed by everybody. We must learn by doing, and a good starting point is to begin from something small that allows us to take pragmatic and concrete decisions. To this aim INSEE-Orange Labs, for example, have worked with old and aggregated data (2007) to avoid confidentiality problems and to help to create that trust.

Related to this, some MNOs called for greater clarity and precision over NSIs needs and the data that may be required on a continuous base for production. However, NSIs find difficult to pinpoint without having accessed the data, hence the importance of developing practical case studies.

- 2) There was a consensus on partnerships outperforming mandatory scenarios. One of the most important aspects is to build trust between the public sector and private companies and the most effective way is to do it within a collaborative environment. Empathy has to lead the manner to approach MNOs. Public interest cannot disregard the investment made by MNOs nor their fair right to take commercial advantage from big data. Efforts have to be made in order to assure a common interest, so that NSIs could also offer to MNOs added value (statistical methodological knowledge, corporate image...).
- 3) A need to find and to agree on the boundaries of markets to avoid a business conflict between public and commercial interests was recognised. Maybe the key is not to define specific domains of customers for NSIs and MNOs but to come to an understanding on the granularity of the information disseminated freely by official statistics. That is the level where the line has to be drawn to find the win-win situation.
- 4) Concerns arise when moving from research to production. This aspect is closely related with the two last points. Cooperation with Orange Lab, for instance, was possible because from the beginning the project was shown as a small and limited research study not production oriented. Moreover, the use of big data was not at that moment at the business core of the company and therefore there was space for a project of that characteristics. The question of moving to industrialised process is totally different, not yet solved and requires a very careful analysis. There are different levels of maturity and the stage of production is seen by the time being in the long term.
- 5) Distributed vs centralized models. Solutions based on the development of open algorithms (OPAL) that could also be used for commercial purposes appear as an interesting candidate. This could be a good approach satisfying NSIs needs for auditable methodology. The bottom line is not to throw data onto algorithms but open algorithms onto undisclosed data.
- 6) Regulation on data protection and relationships with regulatory authorities is a big issue at national and European level. More clarity is needed to define the (new) complex legal framework where a convergence of business sector regulations (that rules the use of information produced by MNOs economic activity), statistics regulations (that gives public statistics producers the right to access data) and data protection regulations (that generally protects citizens' sensitive data) is already evident.
- 7) Perception of the society on the use of these sources. A common communication strategy is needed for avoiding public opinion to reject the statistical use of mobile phone data.

FRIDAY, 23 SEPTEMBER

SESSION III. CORE ISSUES REGARDING ACCESS TO MOBILE PHONE DATA

1. PARTNERSHIPS BETWEEN MOBILE NETWORK OPERATORS AND STATISTICAL OFFICES: OFFICIAL STATISTICS PERSPECTIVE.

INE was in charge of introducing this topic. Firstly, the global framework was addressed, mentioning the UN Fundamental Principles of Official Statistics, UN Recommendations for Access to Data from Private Organizations and the European Statistical Code of Practice. The issue of confidentiality and privacy is on the base of these principles, and they recognise that private companies have to provide data, although a proportionality and fair balance is demanded between information needs and data required. The cost and effort of providing data access must be reasonable compared to the expected public benefits.

Next, INE gave a short description of the ESSnet on Big Data project and pointed out some hints about different characteristics of the MNOs and aspects related to legal requirements, access conditions and data characteristics.

Concerning official statistics position towards the use and integration of big data in statistical production, INE expressed the following:

- 1) Mobile phone data have a clear public interest for official statistics purposes.
- 2) Official statistics producers are aware of and sensible to the complexity of factors involved in the access to data (with the need to figure out how they interrelate among them): specific characteristics of each MNO, unclear legal regulations, different access conditions (in-situ or transmitted, research vs production...), confidentiality and privacy, public perception, collision of interests...
- 3) To tackle these issues and develop case studies, NSIs need preliminary access to data.
- 4) Partnerships / joint ventures are the optimal approach to meet public and private interests.
- 5) Common solutions will help to guarantee the privacy and confidentiality of data, as well as a sustained access to data source to assure statistical production.

Once again, MNOs noted that it was necessary to define what kind of data sets and variables are expected to be collected by the NSIs because on that depends the effort that has to be done to extract and store the data. NSIs replied that fixing a priori the amount, characteristics and frequency of data is not possible, research and case studies are needed in advance. In any case, all participants agree on the relevance of developing a model based on a fair balance of the benefits, needs and costs from both sides.

NSIs also mentioned that data integration was an area where Official Statistics could bring very much knowledge and added value on commercial production, both methodologically and providing linked aggregated data from official statistics and other administrative sources. The role of NSIs as a potential trusted party where to aggregate and integrate data from diverse MNOs to control market share bias was mentioned as an interesting option.

2. PARTNERSHIPS BETWEEN MOBILE NETWORK OPERATORS AND STATISTICAL OFFICES: MNOS PERSPECTIVE.

Telefonica, as a renowned company in European telecoms market and user and developer of big data solutions, presented this issue and gave their MNO perspective.

Telefonica is a data driven company, and big data is one of the strategic lines fixed for the future and to which all the company is committed with. The structure has been changed for this goal (bringing with it also a cultural change) and important acquisitions and investments have been made (staff profile, IT...) to reinforce their capabilities and offer.

At the present, Telefonica is using big data technologies both for improving internal processes and increasing offer to customers.

Next, the company briefly explained their market product, SmartSteps.

SmartSteps uses active and passive events and defines two types of elements, a “stay” when the mobile user is at the same location at least 30 minutes, and a “journey” in case any change of location in less than 30 minutes happens.

The starting and ending location destination are identified and stored. Also, each event is related to a user ID, a time stamp and a geographical location.

First, data is anonymized and put into the system to be treated (data storage and formatted), then data is aggregated per cell (no individual is identifiable) and finally extrapolated by applying algorithms to represent the entire population. A set of socio-demographics variables to evaluate the data and the accuracy for the different studies.

To guarantee privacy and confidentiality, a minimum level of anonymization and aggregation is fixed in 15 users.

In the framework of the ESSnet on Big Data project Telefonica and INE have been in contact to explore the possibilities of accessing mobile phone data. The idea was to develop a case study to investigate, in a specific statistical domain, the feasibility of using this information source in the official statistics production. The research would be concentrate on the study of human mobility using the events generated in the mobile network.

Telefonica’s proposal data access is based on SmartSteps product and would comprise one natural year of data for the whole country. The company would be the owner of the database and would provide supporting infrastructure allowing INE to access the data, build the analytics and export the results.

After debate, the general feeling of participants was that we have a complicated way to go before making a public-private partnership for accessing mobile phone come true. There are many issues to be solved, especially if want to move beyond the field of research and analysis. The question is to start working together in a concrete and small pilot study trying to obtain results that may shed light on a future model for public-private collaboration.

3. TOUR-DE-TABLE AND DISCUSSION: EXCHANGE OF VIEWS ON THIS TOPIC

DG Connect acted as moderator of the tour-de-table. The main topics discussed by the assistants were the following:

- 1) How to break the vicious circle in data access: it is necessary to build detailed case studies and delimit a precise set of data to ask to MNOs....but some kind of data access is needed for setting up these detailed case studies....Although the participants had different opinions, there was a general agreement on giving priority to case/pilot studies.
- 2) Relationships with MNOs could be different depending on their concrete situation and current strategies on big data. NSIs must take into account these different starting points, particularly when trying to solve the question research vs production. An MNO having invested to develop a business line around the statistical exploitation of their data does not face the same situation as an MNO not having this strategy on the core of their business.
- 3) In this sense, it is capital to show that there is no collision of interests between the dissemination of official statistics based on mobile phone data and commercial products and services based on the same data. Not only should there be ample room for both businesses but also they could benefit from synergies since official data are of great complementary value for these commercial solutions. The limits of the market could be associated to frequency and geography granularity, meaning that at certain high level official statistics would provide free data but at further detailed level MNOs would have assured their market. This approach could even promote commerciality. Participants agreed that this issue would have to be worked out in detail.
- 4) Many different aspects and connected among them. We have to identify and understand all these interrelationships. For example, legal environment is still a grey area and MNOs ask for more legal clarity, producing guidelines on data protection.
- 5) A public communication strategy is needed to explain and convince the society of the benefits and guarantees of the use of big data in official statistics. Perceptions will need to be analyzed. Citizens are highly sensitive concerning their personal and private data and their potential use by public administration, especially when these data may come out of the MNOs' information database. A general negative public opinion about MNOs transferring data to NSIs can affect their decision to collaborate.
- 6) It is important to find ways to motivate MNOs cooperation, to offer them incentives for collaboration: improve and increase public image and visibility for their corporate social responsibility, define markets (granularity), return of information, gain new knowledge from official statistics that can be used for commercial products, validation or quality assessment using data from official statistics (e.g. geocoded population registers), overcome market share bias by letting NSIs integrate and aggregate data from diverse MNOs...We also have to be aware that finding motivation is easier for MNOs that have not already invested great resources in a big data strategy and have not developed this competence yet.
- 7) A quality assurance framework for our users is also relevant, and both, NSIs and MNOs should have interest on it. Quality indicators and framework (quality stamp) can be more easily reached by collaboration.

SESSION IV. CONCLUSIONS FOR THE FUTURE

1. JOINT CONCLUSIONS FOR FUTURE DEVELOPMENT / COLLABORATION

INE and Eurostat summarized the main aspects and key ideas discussed during the workshop and thanked the intense and fruitful participation of attendees. They also highlighted the impact of events like this workshop that, from a constructive approach, help the dialogue and understanding of the needs and interests of the public and private sectors in the field of big data.

Eurostat stressed the relevance of the many issues discussed during these two days and the importance of maintaining the bridges open and active for communication. Options will be explored to keep the dialogue and communication between INES and MNOs in order to progress, particularly, in one of the priorities identified repeatedly in different interventions: building new case studies.