ESSnet Big Data II

Grant Agreement Number: 847375-2018-NL-BIGDATA

https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata
https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en

Workpackage L

Preparing Smart Statistics

Deliverable LM3: Final Report

Final version, 31st October 2019

Prepared by:

Natalie Rosenski (Destatis, Germany)

Workpackage Leader:

Natalie Rosenski (Destatis, Germany)
Mail address: natalie.rosenski@destatis.de
Telephone: +49 611 754284

Table of contents

1 Goal of WPL

The aim of workpackage L “Preparing Smart Statistics” (WPL) is to examine the extended use of the Internet of Things (IoT) in order to produce trusted smart statistics for the European Statistical System (ESS). As the range of topics regarding the subject IoT is huge, the goal is to provide an overview of relevant topics for official statistics, to show their variety and to highlight topics that are promising and could be analysed further. Therefore, the availability and accessibility of the different data sources will also be checked.

Furthermore, the goal is to explore how the digital footprints of daily life created by human wearables, city and vehicles sensors and other smart systems could change the way to produce trusted smart statistics taking advantage of societies’ datafication.

The development of trusted smart statistics aims for data that might be pre-processed by the data providers or data sources ready to use for official statistics, which leads to a combination of privately held data with official survey and/or administrative data.

WPL will prepare the ground for future actions on trusted smart statistics by concentrating on the IoT data sources and devices with potential relevance to official statistics, resulting in an overview of the data landscape. The results of WPL should enable future actions to result in proof of concepts and experimental statistics.

All the results of this WP keep in mind that the overall goal is the development of generic solutions and harmonized as well as standardized approaches and recommendations for the ESS.

There are four main topics and therefore tasks conducted in WPL: 1. Smart Farming, 2. Smart Cities, 3. Smart Devices and 4. Smart Traffic.

As the objective of WPL was to give an overview of the 4 topics and to give recommendations for possible follow-up studies, the duration was only 12 months, conducted from November 2018 to October 2019.

There were 12 partners from 11 countries involved in WPL (leader in bold):

- Statistics Austria (STAT, AT)
- The Bulgarian National Statistical Institute (BNSI, BG)
- The State Statistical Office of Berlin-Brandenburg (SSO BB, DE)
- **The Federal Statistical Office of Germany (Destatis, DE)**
- Statistics Finland (FI)
- The National Statistical Insitute of France (INSEE, FR)
- The National Statistical Institute of Italy (ISTAT, IT)
- Statistics Poland (PL)
- Statistics Netherlands (CBS, NL)
- Statistics Norway (NO)
- Office for National Statistics (ONS, UK)
- Statistics Portugal (INE, PT)

Also involved as a subcontractor of BNSI was the company Mimirium Ltd, in the work of task 2, case study 1.

2 Meetings

The kick-off meeting of WPL took place in Wiesbaden, Germany, from 15 to 16 November 2018. All 12 partners were represented in the meeting. Furthermore, the coordinator of the ESSnet Big Data II and two representatives of Eurostat attended as well.

The report on the kick-off meeting (LM1) was sent to the partners of WPL, Eurostat and Statistics Belgium (in order to be uploaded on Wiki) on 31 January 2019 and therefore met the set deadline.

The second and final physical meeting of WPL was hosted by ONS and took place in London, UK, from 3 to 5 June 2019. 11 partners and a representative of Eurostat were present; Statistics Norway could not attend the meeting.

The report on the final meeting (LM2) was sent to the partners of WPL, Eurostat and Statistics Belgium (in order to be uploaded on Wiki) on 23 August 2019 and therefore met the set deadline.

Besides the physical meetings, 6 WebEx meetings were conducted to discuss the state-of-the-art.

The milestones and deliverables of WPL are available via the following link:

https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WPL_Milestones_and_deliverables

3 Content of the four tasks

Details about the four tasks can be found in the four deliverables of this WP.

3.1 Task 1 – Smart Farming

Task 1 was conducted by STAT, Statistics Poland and Destatis. The aim of this task is to explore the possible use of data coming from precision farming in agricultural statistics and to identify potential data sources. In this regard, task 1 provides an overview of the data landscape, that is of smart technologies and agricultural technology providers, and of the data generated by these technologies. Furthermore, it explores the use of smart technologies by farmers on the example of Germany. Finally, it presents a promising use case and gives recommendations for possible follow-up studies.

Generally, data generated by smart farming technologies are very promising for official statistics, but - mainly due to the restricted use of digital technologies by the farmers - it might be more interesting for official statistics in the long run. Nevertheless, data sources explored in the European project "Big Data Grapes" might also be interesting for official statistics and could be examined further in a follow-up study.

3.2 Task 2 – Smart Cities

Task 2 was conducted by BNSI, INSEE, ISTAT, SSO BB and Destatis. This task aims at investigating the potential of IoT tools (such as smart sensors) in order to produce smart statistics related to the topic Smart Cities.

In case study 1, BNSI explored if new technological solutions of Smart Cities, such as Blockchains or LoRaWAN (a media access control protocol for wide area networks), can be used for the purposes of official statistics. In this regard, information about problem areas for disabled people was collected via an app, that was developed for this case study, and via GPS and additionally installed devices in the city. The case study shows that the used tools allow to collect data that previously couldn't be collected and at the same time secure the privacy and confidentiality of the respondents. This was

achieved by storing the gathered data only locally in an encrypted model on the users' devices and by only transmitting anonymized and aggregated data to the National Statistical Institutes (NSIs).

It is highly recommended to establish not only an ESS interest group that builds knowledge and skills about the chances of the technologies used here, but also to generate more use cases based on the technologies explored here and to test the use of these solutions for all member states.

In case study 2, ISTAT, SSO BB and Destatis researched the outcomes of the 12 projects that were funded by the Horizon 2020 Programme of the European Commission. Furthermore, the cities Berlin and Darmstadt were explored regarding the smart technologies they implemented and how the generated data might be of use for official statistics. The aim of this case study is to give a first overview of the different 'smart solutions' that are partly still developed and partly already implemented in cities across Europe and to highlight the most promising solutions for official statistics.

Although there seem to be a lot of valuable data generated by Smart Cities, the availability and access to it is generally rather poor at the moment. Nevertheless, for a follow-up study it is recommended to further explore the city Darmstadt, especially as there is already some data available as open data that was collected via sensors on traffic lights. Another use case could explore the data platforms that combine all data that is generated by different technological solutions in Smart Cities. The third use case could further investigate so-called smart lamp posts that capture different kind of data, such as data about vehicle traffic, air pollution and water level.

In case study 3, INSEE and ISTAT analysed the correlation between socioeconomic characteristics of people and the level of pollution in the cities Nice and Rome. Especially in Nice, a successful cooperation with partners that provide and use raw open data was established. In Nice, people of higher socio-economic status are significantly less exposed to air pollution than people of medium and low socio-economic status. This was probably the most advanced case study of task 2, combining open source data, used by a Smart City, with official data.

Possible follow-up studies could include on the one hand an incitation to other countries to follow the example of INSEE and ISTAT and on the other hand methodological developments in order to improve the quality of the estimation. Also, a deeper collaboration with smart cities is foreseen, in order to better understand the reasons for pollution by analyzing traffic data collected by the city.

3.3 Task 3 – Smart Devices

This task was conducted by CBS, INE and ISTAT. The aim of this task was to conduct a first study on smart devices from the viewpoint of official statistics. It was meant to be preparatory for future projects on trusted smart statistics that will be able to go deeper into the most promising use cases identified in this task. Task 3 provides a comprehensive list of different types of smart devices that are categorized as follows: smartphones, smart home devices, smart devices for health and fitness, smart devices for mobility, smart devices for travel and other smart devices. The devices are described regarding their purpose for the users of the specific device, their distribution and also concerning their potential use for official statistics.

Generally speaking, data from any smart device or citizen science project is very interesting for official statistics. However, preconditions are the type of data access (if accessible at all), the kind of data it concerns, the maturity of the devices and the transferability to the ESS. In this regard, three use cases were identified: 1. The realisation of a citizen science project in order to examine voluntary public participation for statistical purposes, 2. the development of an app to deduct

people's mobility pattern using machine learning techniques and 3. the analysis of travel patterns based on smart travel cards.

3.4 Task 4 – Smart Traffic

Task 4 was conducted by Statistics Finland, Statistics Norway and ONS. It contains two subtasks exploring new data sources that can be used to analyse traffic for official statistics.

In subtask 4.1, Statistics Finland and ONS explored the use of traffic inductive loop sensors for monitoring traffic in order to use it for economic estimates. Although there are some data quality issues, traffic loop data has many advantages, especially the availability of time series data often going back to times prior to the 2008 financial crisis. ONS opts for creating a set of indicators that allow early identification of large economic changes in trade, whereas Statistics Finland derives truck traffic growth rates to produce early estimates of GDP. These new products based on traffic loop data are able to give policymakers early warning of changes in the economy, although care must be taken when using them and it's important not to conflate them with regular official estimates of GDP and trade. Finland and the UK published their research on their websites already and will continue to improve the products. They highly recommend that other NSIs follow their example.

In subtask 4.2, Statistics Norway explored promising new data sources for producing statistics on road freight transport. From all the data collected nowadays by modern trucks, three data sources were identified as particularly promising: Build-in sensor data, GPS-data and tachograph data. If the data would be used by official statistics, the response burden in the freight transport survey would decrease significantly. In order to use the data, either a uniform interface for data delivery needs to be built or alternatively third-party GPS-data could be used or integrated smart tachograph data could be utilized as position data.

The data source is evaluated as interesting for the whole ESS, as information requirements, quality issues and technologies on sensor data in trucks apparently are very alike throughout European countries.

4 Conclusion

In WPL, a wide variety of topics concerning the subject IoT have been examined and many promising data sources have been identified. The results show that it might be rewarding to dig deeper in follow-up studies. However, generally the data landscape is still in a development state. Data access and availabilty are often an issue. Even though merging of sensor data with geolocalized official data has been tested succesfully and some data sources are already used for estimates and have been published as research papers, the data format and the transferability to the ESS still has to be examined in many cases. Also, data protection is an important subject to take care of by official statistics, but there are quite promising trends leading to open data, which is usually anonymized and aggregated. Also, the confidentiality of data could be protected by using new technologies and procedures such as cryptography technologies or blockchains. Certainly, these technologies have to be further investigated. In addition, the quality of data gathered by smart systems and their representativity should be examined carefully as well.

Altogether the use of data generated by IoT tools offers many possibilities to enrich offical statistics, but both using new data sources and adopting/changing the statistical production processes of the NSIs to involve those have to be explored in further studies.