Federal Statistical Office of Germany
B 203/12203000
Natalie Rosenski

Wiesbaden, 31 January 2019

Report

on the kick-off meeting of WPL of the ESSnet Big Data II,

at Destatis, in Wiesbaden,

on 15/16 November 2018

Participants

Country / Organisation	**Name**
AT	Gregor de Cillia
BG	Valentin Chavdarov
DE / Destatis	Jasmin Arnold
DE / Statistik BBB	Holger Leerhoff
FI	Henri Luomaranta
FR	François Sémécurbe
IT	Massimo De Cubellis
NL	Olav ten Bosch
NL	Peter Struijs
NO	Johan Fosen
PL	Artur Laczynski
PT	Sónia Quaresma
UK	Alessandra Sozzi
Eurostat	Albrecht Wirthmann
Eurostat	Martina Hahn
DE / Destatis	Natalie Rosenski
DE / Destatis	Clara Schartner (1st day)
DE / Destatis	Thomas Fuchs (2nd day)

0. Welcome and Introduction to the ESSnet Big Data II – WPL

Natalie Rosenski welcomed the participants to the kick-off meeting of WPL of the ESSnet Big Data II.

Albrecht Wirthmann gave an overview of the topic "smart statistics" and the expectations of Eurostat regarding the activities of WPL. Peter Struijs put WPL in the context of the ESSnet and gives some background on WPL.

Both reminded the participants of WPL to have in mind the follow-up already during the current activities and that information on these activities should be reported early (to Eurostat) in order to prepare for future grants and future proof-of-concepts.

In WPL, cross-cutting issues, which are quality and metadata framework as well as the identification and definition of the target architecture and the required IT capabilities, should also be examined. Skills and competences necessary to fulfill the identified tasks should also be considered. The way forward should be outlined explicitly in a roadmap.

1. Task 1 – Smart Farming (AT, DE, PL)

Regarding task 1 a joint presentation from Austria and Germany on the topic of Smart Farming took place. The speakers dealt with the issues of deployment of new digital technologies, challenges, implications for statistics and potential data suppliers. The presentation started with a brief overview of the use of digital technologies in the agricultural sector. It became clear that the different technologies are used to different extents. While, for example, satellite-based systems are already used by more than half of all farmers, sensor technology, drones and robots (with the exception of livestock farming) play hardly a role. This circumstance should be taken into account.

The next part of the presentation dealt with the challenges of smart farming. Particularly noteworthy are the low network coverage in rural areas and the lacking compatibility between agricultural machines as well as software solutions. In addition to these two points the subjects data protection and media competence of the farmers are also important.

Destatis is aiming in the long term to develop a distinct data output interface for agricultural statistics via a platform. However, this goal cannot be achieved within the framework of the ESSnet. This means that in task 1 only the first steps on the way can be taken.

Finally, Austria presented its plan to create a catalogue of potential data suppliers. Those data suppliers could be farmers, unions or tech companies who produce smart devices. The catalogue should achieve two goals. First, it can be used to link data requirements to corresponding sources in the future. Second and more importantly, the catalogue will provide a clear idea on the data-landscape as well as the most commonly used devices by farmers. It is planned to also talk about a possible cooperation in the future during the communication with the potential suppliers.

Discussion:

- Task 1 should initially concentrate on the activites in the countries Austria, Germany and Poland, but consider not only a comparability between the countries, but also share the results or overviews with NSIs that are not involved in this task.
- At a later stage, a survey with other NSIs could be considered.
- One important issue to be answered by this task is the question, how the data can be acquired. Therefore relationships should not only be built with farmers, but also with potential data suppliers such as manufacturers or service providers.

- The work of the Directorate-General for Agriculture and Rural Development (DG AGRI) should be consulted for this task.
- Another question to be answered is about the prerequisites to use the data, e.g. technical questions and questions of data access.
- The result should be a landscape of different farming systems/sensors and different types of actors (e.g. farmers, machine producers, software engineers) in regards to different fields of farming and related data availability.
- Also data flows and data ownership should be looked at.
- The goal of this task is to determine a fitting strategy and to undertake a proof-of-concept in a follow-up project.

2. Task 2 – The Use of IoT for Smart Cities (Berlin, BG, DE, FR, IT, UK)

Regarding task 2, case study 1, the Bulgarian NSI will explore opportunities to produce meaningful statistics from the data coming out of sensors and mobile applications in the city of Varna.

The city Varna was chosen, because a good part of the technical infrastructure for submitting data comes out from sensors and mobile applications to a server that is in place. We have to decide on what to measure by means of sensors and mobile applications attached to existing infrastructure: traffic intensity, people movements across the city, quality of air, parking space, gaps in the public transports, etc. The list of possible use cases is quite long. This requires development of a methodology to transform the huge flow of raw data into a relatively small number of meaningful indicators, helpful for both citizens and municipal administration. This is the "methodological" part of producing smart statistics on the city level.

We will share the experience with the city of Varna with other cities in Bulgaria. Hopefully we will produce smart city statistics on everyday life.

Regarding task 2, case study 2 on *Smart Cities and Communities lighthouse projects* a joint presentation by Destatis, Istat and Statistik Berlin-Brandenburg has been illustrated.

In the presentation, the 12 Smart Cities and Communities lighthouse projects funded by the European Commission have been listed and some possible examples have been illustrated.

These projects aim at demonstrating how to combine ICT, e-mobility and energy solutions to design smart and resilient cities for citizens and companies, in order to (i) improve quality of life of citizens, (ii) reduce the environmental impact of activities and (iii) create a stimulating environment for sustainable economic development.

The projects try to use integrated approaches, that means scalable and replicable solutions. An objective of this use case could be to analyze, among the different projects, if there are some common aspects in terms of data, devices used and methodologies applied.

In particular, in this case study, these projects will be examined with a focus on the data generated by the smart cities involved in the projects and on how these data could be used to produce official statistics.

Regarding task 2, case study 3 on *Air Pollution* there has been two presentations: one by Istat and one by Insee.

The presentation of Istat was divided in two parts: in the first one, the traditional official survey on air quality has been briefly illustrated and in the second part, a design idea to study the socio-economic characteristics of people exposed to pollution with the aid of smart sensors, has been presented.

About the traditional survey, Istat provides annual statistics on air pollution for the 120 main cities of the country, in compliance with the European regulation. The data are collected by the statistical offices inside the cities, and are provided by the Regional Agency for Environmental Protection (Arpa).

The micro-data refer to some polluting substance, such as PM10, PM2.5 (Particulate Matter), NO2 (Nitrogen dioxide) and O3 (ozone), detected by the devices located inside the main Italian cities.

Passing to the design idea, the goal is to provide more detailed statistics in spatial and temporal terms with the aim of studying the socio-economic characteristics of the population exposed to pollution.

To reach this goal, Istat would use the same micro-data used by the statistical office on pollution, enriched with the implementation of a geo-spatial model to represent a more detailed map of pollution point by point, street by street (the idea is a sort of Google Traffic map applied to pollution). There are several models that are able to estimate how the pollution detected by the sensor is distributed on the geographical space. This models use as input the pollution data, the weather data (temperature, wind, sunning/raining, etc.), chemical data refer to the characteristics of each substance measured by the sensors, and provide as outcome a pollution chromatic map. The goal is to link these micro-data with the census data and in particular the official data of dwellings and resident population, at very low level to study the socio-economic characteristics of the population exposed to pollution.

Insee has presented its project of collaboration with the smart city innovation center from the Nice metropolis. The metropolis has installed many sensors that measure pollution data. It allows to have access to very accurate information, without having to rely on a model. The aim of the collaboration would be to relate these pollution data with geo-localized socio-economic data produced by Insee, in order to better understand the characteristics of population exposed to a high level of pollution. If the implementation of a pilot project in the Nice metropolis succeeds, this could act as Proof of Concept in order to help Insee to have access to the data of other municipalities equipped with sensors and to reproduce this study. The long term objective is to be able to give local public actors insights into the urban quality of life in different French cities.

Concretely, in order to make the link between Insee's data bases and the sensors' data, Insee and the smart city innovation center are planning to hire an intern who will be jointly supervised by both parts. This internship should begin in May 2019.

The two NSIs have the same goal: to study the socio-economic characteristics of people exposed to pollution with the aid of smart sensors.

In this case study , the first step for Istat and Insee is to identify similarities and differences of their design ideas; the proposal is to define in more detailed way the two experimentations in order to compare them in terms of data, methodologies and techniques.

Discussion:

- A definition of the 'Internet of Things' (IoT) in coordination with the definition of smart devices (task 3) should be included.

- The purpose of the actions in relation to the European Statistical System should be clear.
- Standardization is important.
- Although the case studies are independent from each other and relate to specific cities, connections between the use cases and therefore a communication between the use cases should be ensured.
- One of the goals is to prepare the way forward and to include an European perspective.
- Creation of a list of characteristics or criteria for which the analysis of the Horizon 2020 projects can be done.

3. Task 3 – Smart Devices (IT, NL, PT)

Regarding task 3, on smart devices, there was a joint presentation from the Netherlands, Portugal and Italy about the first ideas on the subject. The presentation started with a review of the general definition of smart devices on Wikipedia. One of the observations was that the world of internetworking protocols for IoT and smart devices is still developing. New protocols, such as LORwan and ANT+ for different use cases were added recently and new protocols are being developed regularly. This is important to take into consideration in the exploration of smart devices. In addition there was some discussion and agreement on the *interactive* and *autonomous* characteristics of smart devices. All-in all, the participants agreed on the definition presented as a first starting point for the work being done.

Many examples of smart devices were presented, both from home equipment, health, safety, transport as well as some weird ones. The importance of the introduction of the 5G network was explained and the ideas on a statistical view of the subject were presented.

There were a number of examples of projects presented, such as a students' study on health from Statistics Netherlands, a project to derive data for quantifying the Dutch eco-system from the running app Strava and citizen science on air quality. It was noticed that citizen science is a subject that relates to smart devices and should be taken in consideration. Some dedicated search engines for searching the IoT world were mentioned in the presentation and four possible models for data access on smart devices were presented as well as some real life ideas how this could work in practice. The role of public administrations, the citizen science concept and aspects of representativity / quality were mentioned and finally a wrap up of the plans for the project was presented.

Discussion:

- The work of task 3 should start with a definition of smart devices and a distinction from other tasks. This input is of interest to the other tasks as well.
- The notion of smart systems as defined in the reference architecture presented by Eurostat on the DGINS (Directors General of the National Statistical Institutes) conference should be taken into consideration.
- This task aims to include data that citizens generate with their private smart devices and that they then share in open/public platforms, bringing awareness to the citizen science phenomenon.
- A mechanism should be identified how this kind of data could be used for official statistics.
- A methodological issue could be that a device might change the behavior of people.
- The assessment of the use of a device and under which conditions NSIs can use the data is important.

4. Task 4 – Smart Traffic (FI, NO, UK)

Regarding task 4, subtask 4.1, Statistics Finland presented the ideas on producing nowcasts, or early estimates of economic indicators, based on traffic loops data. This work builds on experiences and results gained in the ESSnet on Big Data I (2016-2018), where several nowcasting methods were tested. They were based on machine learning algorithms that extracted information from the combination of traffic loops and firm level disaggregated information. The subtask 4.1. focuses on using the traffic loop data in a realistic setting, exploiting its timeliness (real-time) and tries to establish a system that is able to extract and analyze the data and possibly produce an automated statistical result. The Slovenian statistical office (SURS) and the statistical office of the UK (ONS) are simultaneously developing similar tests.

For Subtask 4.2, Statistics Norway presented the planned work which is to make an inventory of the smart meters inside trucks. There are several smart meters inside these trucks and it is interesting to get an overview of the data collected from the meters. The inventory will also look at who the data owners are and what possibilities there are to get some of these data to Statistics Norway. The latter involves contacts with institutions collecting sensor data. Finally, the inventory will look at which potential use can be anticipated regarding the production systems of official statistics.

The inventory will be a base for future data collection by Statistics Norway and of use in official statistics production.

A relevant Norwegian project is LIMCO: a large research projects (Institute of transport economics in cooperation with industrial partners) aiming at transport planning and optimization and sustainable business models.

Discussion:

- Subtask 4.2 is quite different from subtask 4.1, but any insight being relevant for subtask 4.1 will be shared with subtask 4.2 and vice versa.
- For subtask 4.1, a comparison to the work of the Dutch and Slovenian colleagues is possible.
- Results of 4.1 could be used to communicate results and promote them.
- National accounts should be involved from the start in subtask 4.1.
- Subtask 4.2 should give an overview of the landscape of sensor- and other transport data.
- Emphasis on the relevance for other NSIs.

- Analysis codes/algorithms should be shared on the Wiki of the ESSnet.
- Generally, in Nordic countries there are special arrangements between the offices to exchange data.
- The follow-up and recommendation for an action in the next phase should be considered in this task as well.

5. Cross-cutting issues

Within WPL:

- It is recommended that all participants of WPL read the drafts of all tasks to recognize cross-cutting issues.
- The definition, landscape and classification of smart devices (task 3) are very important for the other tasks as well.
- If the other tasks characterize their sources as well, the tasks could all be connected and smart devices can make sure that everything is thought of.
- Task 2 should paint a picture of the connection between the devices and how they work together to construct an overall picture of IoT.
- The overall goal is a smart system of statistics to combine data from different countries and domains. Therefore it is necessary to structure that (e.g. common reference frame, common information flow).

With other WPs:

- There are a lot of connections to the other WPs, e.g. WP F, H, I, J.
- An intensive communication with all WPs is not possible, but the occasional participation of the leader of WPL at the web meetings of the pilot and implementation track seems reasonable.
- An exchange with the other WPs is important in order to avoid double work.
- Informal ways of keeping informed should be used as well.

6. Action plan and schedule

Evaluation remark of Eurostat:

- The evaluation remark of Eurostat regarding the clarification of proof-of-concepts was solved. The participants agreed on the fact that proof-of-concepts are prepared in this WP, but that they are not conducted yet.
- Process steps regarding to Eurostat: Scoping → Proof-of-concept → Feasibility (Prototype) → Implementation

Schedule:

- It was agreed to have WebEx meetings every two months; an outline or structure should be sent one week ahead of the meeting.
- The final report will be sent to the review board so that the draft should be sent to them two months in advance.
- The 2nd meeting is proposed to be in June, the report will be finished in August 2019.
- Alessandra Sozzi will check at ONS if a meeting in their premises is possible.

Outline of the reports:

- The reports on the meetings should be extended minutes.
- It was agreed that Destatis will provide a draft of minutes within a week and that the participants (especially the presenters) will add content to the minutes.
- The structure of the four deliverables should be exchanged at an early stage, in order to have similar structured reports (e.g. on the Wiki) and to make a comparison more easy.

Miscellaneous:

- Martin van Sebille (Netherlands) is organizing the WebEx meetings if a request is sent to him in due time.
- Peter Struijs will not attend the WebEx meetings, but Albrecht Wirthmann will.
- Regarding the sharing of documents it should be avoided that documents need to be shared at different portals.
- Wiki is the reference place for all the material and can also be used to show which statistics are developed (it is used in particular for collaboration purposes).
- There is an external and internal Wiki (restricted area according to log-in).
- For ESSnets, it is obligatory to put results on Cross portal (there are guidelines available for ESSnets how to use Cross portal).
- Marc Debusschere (Belgium) is responsible for Wiki and Cross portal.