# Statistics Netherlands

**Minutes**
**ESSnet Big Data II**

| | |
|---|---|
| to | Co-ordination Group ESSnet Big Data |
| cc | |
| from | Martin van Sebille |

| | |
|---|---|
| subject | Final minutes CG meeting 8 (Pilots Track) of ESSnet BD II by WebEx 17 April 2019 |
| | 2019-04-29 |

*Participants:*

| | | | |
|---|---|---|---|
| Peter Struijs (Chairman) | √ | Marek Morze (PL WPH) | √ |
| Martin van Sebille (Secretary) | √ | David Salgado (PL WPI) | √ |
| Marc Debusschere (Deputy WPA) | | Marek Cierpiał-Wolan (PL WPJ) | √ |
| Tomaž Špeh (PL WPB) | | Alexander Kowarik (PL WPK) | √ |
| Galya Stateva (PL WPC) | | Natalie Rosenski (PL WPL) | |
| Arko Kesküla (PL WPD) | | Albrecht Wirthmann (Eurostat) | √ |
| Anke Consten (PL WPE) | | Lilli Japec (Chair Review Board) | |
| Monica Scannapieco (PL WPF) | √ | Anders Holmberg (Review Board) | √ |
| Johan Fosen (PL WPG) | √ | | |

*Replacements/Guests:*

| | | | |
|---|---|---|---|
| Kostas Giannakouris(Eurostat) | √ | Fabio Ricciato (Eurostat) | √ |

## 1. Opening and agenda

Peter welcomes everybody to the eighth meeting of the Coordination Group of the ESSnet Big Data II (CG). This meeting is the second meeting of the Pilot Track.

Marc and Lilli cannot attend the meeting; they sent notifications. Anders will join the meeting somewhat later.

There are no agenda changes. Peter remarks that if a specific subject is relevant for discussion in this group as a whole please inform Peter and the request will be accommodated.

Another remark concerns the signing of the Grant Agreement (GA). This was formally done last Friday.

## 2. WPK: state of affairs and matters to be coordinated

The planning of the mid-term meeting for the pilots track is 11-12 December in Vienna. On 9-10 December the mid-term meeting for the implementation track

is planned, also in Vienna. The order of these meetings is not definite yet, this depends on whether the WP leaders can attend.

The face-to-face meeting of WPK will take place in Lisbon on 13-14 May and the major topic is the typification structure, for which a matrix is being drafted. The matrix is not yet ready to be put on the wiki. When the draft is ready WPK will try to fill it out with the feedback from partners.

The matrix has two dimensions: data challenges (e.g. on the roadmap to products, privacy and access, metadata) and data types.

Fabio wonders if, considering access, this includes the type of data per sector, such as transport or energy. And are stakeholders on national and transnational level, legislations and other authorities also taken into account? The reason why Fabio brings this up, is to be aware of the fact that legal aspects not only concern privacy but also the national regulations per sector. The approach needed can differ per country.

In principle Alex is aware of this, and the matrix can accommodate this, but not at a detailed level. It also depends on the input from the partners. This will also evolve further.

WPK has started with the quality framework and the typology has links with the methodological framework.

3. **WPG to WPJ: state of affairs**

*WPG: Johan*

For the first deliverable in October a general chapter will be written. The other chapters will be more specific for each country (partner) in this WP. An inventory on the availability of financial data among the partners is being composed. Peter suggests that the inventory should not only be done for the partners within this WP but also for the other partners in this project. For instance, at the ESS Big Data Task Force meeting in Paris earlier this month, Denmark mentioned that they may also use certain financial data for statistical purposes. Peter will send the list of partners to Johan.

Concerning the shared economy platform the WP has finished with the definitions. Parallel there is contact with Eurostat experts on this topic.

The face-to-face meeting will take place on 15 and 16 May in Oslo.

*WPH: Marek Morze*

As mention during the previous CG meeting, all partners of WPH have adopted the methodological framework proposed by the Polish team (it is on the wiki). It contains a table with the detailed description of subsequent stages of satellite imagery processing.  This was accepted as common tool to describe activities conducted in the case studies. After fixing review dates the plan of activities should be prepared to produce the first deliverable (the midterm report) in time. This plan will be a subject of discussion during the second WPH WebEx meeting in the beginning of May.

According to this plan WPH should finish the following activities (pre-works stage) until the end of April:

- define definitions of possible statistical products
- finish literature review (a few literature items have been sent to Alex)
- finish data source review (EO data and administrative data)
- finish the software review (open source and commercial)

During May WPH should:

- prepare and describe the test site
- finish collection of EO data and administrative data
- prepare analysis of metadata, IT infrastructure and quality of collected data

From June to the beginning of September WPH plans to describe the data pre-processing stage and segmentation and classification methods.

Each country (partner) has to plan their case study with its own specificity. Important for this is the forthcoming face-to-face meeting to be held at the regional office of Marek in June. The dates are not fixed yet.

*WPI: David*

The WP works on parallel research tracks. The MNO Telefonica sent an offer for providing access to data and the offer must be studied very closely. The focus is on building a partnership collaboration and not a client-provider relationship. The Netherlands is in contact with T-Mobile. This fruitful collaboration (see below) offers the opportunity both from a methodological/technical perspective and also regarding the access to data: the aim is to understand what the exact interest is of the MNOs and the subtleties around the private-public collaboration.

The simulation track is making good progress, with producing some first simple simulated data.

On the track of methodology:

- The Dutch team made a manuscript about the initial general framework for the geolocation of network events. This is the first step to identify the datasets needed for the upper-level stages of statistical analyses to produce official statistics.
- In the last NTTS progress was reported about different hierarchical models to connect these datasets with the target populations under study.
- Work has been done on inference and Bayesian modelling of the whole end-to-end process from network events to final estimates, comprising some quality indicators.
- Just recently a contribution in the form of a poster (together with Eurostat) was accepted for the NetMob conference in July in Oxford. This might lead to more interaction with both MNOs and academia.
- The next face-to-face meeting is on the 19th and 20th of June at ISTAT in Rome. Not the whole WP will be present, as meetings are devoted to different specialized topics. This one is about geolocation. Also Eurostat will be present.

Peter remarks that in order to protect business interests, software may be made public with a delay. He also mentions that at the Task Force meeting in Paris earlier this month, the production framework with mobile network data was mentioned as one of the three initial subjects to be covered by the future

TSS centre, the other two being OJV (online job vacancies) and smart surveys. Concerning smart surveys, a call was published last week.

*WPJ: Marek Cierpiał-Wolan*

The first deliverable (webscraping methods) is finished and for feedback sent to the Review Board. (*Note: At the meeting it appeared that it had not reached the Review Board, there had been some misunderstanding. This was solved immediately after the meeting.*) Maybe it is a bit too early because during the project more experience will be gained.

Further as a contribution to the report a short document is being prepared regarding big data sources: Each country is analysing data sources, the demands and data breakdowns, and the methods how to get the data (also looking at the long term availability). This is the starting point for the next task (1c), taxonomy.

The face-to-face meeting will be held in June in Poland (dates to be decided later).

Peter wonders if the webscraping and other methods can be reused in similar situations, or whether certain types of sources are required as a minimum. According to Marek the aim is to have a universal solution based upon the webscraping methods and not to different sources of data. The developed methods are reusable within the WP but he will check this at other partners.

Monica has developed within WPF an architecture model on logical level, comprising a description of the different business functions and the lifecycle of processes in which big data is exploited. The architecture is meant to be a common reference for all WPs, although certain aspects can be further worked out depending on the needs of the WPs, e.g. concerning webscraping. So it is important for all WPs to give feedback to WPF for the model itself to optimize it.

More in general Peter mentions that if you were too optimistic about the deadline of a deliverable at the beginning, or if you encounter issues, please communicate this early on. If needed, deadlines can be adjusted if agreed with Eurostat.


4. **Matters to be coordinated with WPF**

WPF has two approaches. The first one is top-down. WPF is working on definition and designing the business layer. It will be put on the wiki and the first draft about business functions and process lifecycles has already been uploaded. Now WPF is working on stakeholders and principles. A draft version will be ready mid-May and put on the wiki. This will also be the input for the face-to-face meeting at the end of May (28-29) in Rome.

The second approach is bottom-up. Each WP leader received the questionnaire. The feedback is helpful so it is known which solutions are adopted. The top-down design will be adjusted with the feedback received with respect to the proposal. Furthermore, WPF has started a discussion about the communication structure within the project. It is thought to be important to provide as much information as possible to all WPs. WPF is thinking about how to do that and meetings like this are also helpful.

### 5. Any remaining issues and closing

There were some questions concerning the signing and status of the GA. Kostas emphasizes that the GA is legally valid. So if you want to make arrangements with data holders, fix financial or administrative matters as specified in the GA, please go ahead, as it is now covered.

Marek C sent the WPJ paper also to Austria and asks if Alex has any comments. Alex had a look at it, but not with the intention to give comments. Right now, he has no specific comments.

According to Peter, at the Task Force in Paris, DG Connect gave a presentation on legislation concerning Open Data which may be relevant for all of us who need access to data from private partners. Providing access may imply costs for them. There will be calls from the European Commission for organizations to provide "high value datasets", for which a high budget is available. This could apply to, for instance, MNOs that want to prepare data access for NSIs. There is a procedure for designating "high value datasets", after which a call may be issued to which MNOs may respond. Albrecht confirms this. This is quite new for Eurostat, namely indirectly financing the production of statistical data from new data sources involving third parties. Eurostat will follow this up more closely and inform the WP leaders where relevant. It is a positive development.

### Meetings Tracks and Co-ordination Group ESSnet Big Data till end 2019

| Date | Time (CET) | Issues | How/Where |
|---|---|---|---|
| 2019-05-22 | 10:30 – 12:00 hr* | Regular CG meeting (incl. resources spent) | WebEx meeting |
| 2019-06-05 | 10:30 – 12:00 hr* | Track CG Implementation | WebEx meeting |
| 2019-06-19 | 10:30 – 12:00 hr* | Track CG Pilots | WebEx meeting |
| 2019-07-24 | 10:30 – 12:00 hr* | Regular CG meeting (incl. resources spent) | WebEx meeting |
| 2019-09-18 | 10:30 – 12:00 hr* | Regular CG meeting (incl. resources spent) | WebEx meeting |
| 2019-10-09 | 10:30 – 12:00 hr* | Track CG Implementation | WebEx meeting |
| 2019-10-23 | 10:30 – 12:00 hr* | Track CG Pilots | WebEx meeting |
| 2019-11-20 | 10:30 – 12:00 hr* | Regular CG meeting (incl. resources spent) | WebEx meeting |
| 2019-12-04 | 10:30 – 12:00 hr* | Track CG Implementation | WebEx meeting |
| 2019-12-18 | 10:30 – 12:00 hr* | Track CG Pilots | WebEx meeting |

*connecting from 10:15 onwards all WebEx meetings