



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

subject Final minutes CG meeting 5 (Pilots Track) of ESSnet BD II by WebEx 27 February 2019  
2019-03-11

*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 Keskülä (PL WPD)		Albrecht Wirthmann (Eurostat)	√
Anke Consten (PL WPE)		Lilli Japec (Chair Review Board)	√
Monica Scannapieco (PL WPF and co-chair)	√	Anders Holmberg (Review Board)	
Johan Fosen (PL WPG)	√		

*Replacements/Guests:*

Fernando Reis (Eurostat)	√		
--------------------------	---	--	--

## 1. Opening and agenda

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

Marc will first tell something about the state of affairs of the wiki and Fernando Reis (Eurostat) was last Friday invited to tell about the BDTI in this meeting also.

Marc has a few remarks about the state of communication. The general structure is online. Underneath red links there are no documents and under the blue links there are.

For each WP there is a brief introduction on the wiki of that WP. If the WP leaders want to add or change something they can do it themselves or ask Marc. Each WP entry page has a link at the top and bottom for more information about the WP. Essential information and the list of deliverables and milestones with the dates are uploaded. Please check if the dates are correct. Also all the minutes of the general meetings are on the wiki and this morning a new page for Alex Kowarik (WPK) was made with a template

providing information about the literature (downloadable). This can be found by clicking on the main page at \*Aids and Tools\*.

The best procedure to inform others is not by sending around documents but to put the documents online and distribute the link of the wiki to access the information.

A general remark by Marc which comes back at every meeting and each work package: Just upload any information you have yourself on the confluence part or on the public wiki (preferably) or ask Marc to do this for you.

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

WPK has finalized the template for the literature review and it is ready to be reviewed and extended. The template is a Word file and downloadable from the wiki. Please complete this template with information from your WP. Send the filled in template to Alex and Jacek or at least provide the bibliographic information (or the paper itself); the email addresses can be found on the wiki and in the Word file itself.

Concerning the quality aspects, there is not much information yet. The deliverable is due in month 9 of the project, July 2019. The aim is to have a quality guideline to share. WPK is working on the structure of it and the link will be on the confluence page. In time, when information is mature, it will move to the wiki.

The document will be split up in sections (e.g. introduction, data source, throughput phase, how to use data sources in combination with other data sources). Classes of big data will be distinguished. The details have not been worked out yet.

Peter asks about the plans for tuning the quality aspects with international quality documents in this area. WPK is looking at the different working groups within the ESSnet and takes into consideration the guidelines of the UNECE and how they relate to the guidelines of WPK. After the face-to-face meeting in May there will be a first raw draft and then other working groups will be contacted for comments and additional information.

## **3. Use of BDTI**

Fernando Reis (Eurostat) is asked to give an update of the BDTI.

The BDTI will be launched in a few months but is already useable for the ESSnet. The information document about BDTI will be sent to all WP leaders (and put on the wiki by Marc). After reading WP leaders may contact Fernando for a meeting if necessary. The first meeting is the 28<sup>th</sup> of February with WPE (AIS data). WPI is in direct contact with Fernando about the BDTI and also WPB is involved, as Cedefop has made a request to use BDTI. Cedefop isn't going to set up a datalab itself soon.

The BDTI stalls the received data in the cloud. One cannot get to the data 24/7, to access it this has to be scheduled.

Please read the minutes of the fourth CG meeting (22<sup>nd</sup> of February) for more information about BDTI.

#### **4. WPG to WPJ: state of affairs**

*WPG: Johan*

The WP is working on two lines: financial transactions and shared economy platforms.

First line: All WP members are in contact with national banks. Norway has already the disposal over financial transaction (on a single day).

Second line: WPG is in close contact with Eurostat. The initial aim was to try to get platform data and to find out if this is shareable or not and how the (meta) data is on these platforms. However this seems not be fruitful and in consultation with Eurostat we now change the focus to see how we can use financial transaction data also for getting informed on the shared economy. Today is an extra WebEx meeting on how to progress this task.

So this approach will lead to a more integrated approach between the two tasks (no pipeline approach).

Peter remarks that we will not change the agreement at this stage and that we will deal with the new approach though the current (draft) agreement. Peter also remarks that this may be an issue for the ESS BD Steering Group. What direction do we prioritize? Later this year funding will be available for a new action on TSS.

*WPH: Marek Morze*

Nine partners have prepared their own study case. The range is quite wide.

In the Vienna meeting the use of common standard tools was discussed. Marek proposes that his team will provide such a tool and share it with the others for feedback. The methodological framework, to be applied by all partners, describes four stages of data processing, statistical work and other additional activities, methodology and IT. The framework is the subject of discussion at the next WebEx (28<sup>th</sup> of February).

Until now the activities done concerned reviewing literature, data access, software (open and professional). Each WPH partner has to determine the definitions of possible statistical products of his case study. The results of the review process will be discussed in the WebEx meeting.

Peter praises the initiative taken and remarks that the common methodological frame is very much connected to the work of WPF and WPK. He likes to see that these are coherent and consistent. Please inform Monica (WPF) and Alex (WPK) as early as possible and please consider inviting them where relevant to WebEx meetings as well.

Monica mentions that someone of WPH can also participate in WebEx meeting of WPF to be aware of the architectural activity in WPF.

The first version is a Polish version and then it will be discussed within the WP. If it is more mature it will be uploaded on the wiki. Marek will send the draft version already to WPK and WPF.

*WPI: David*

It was a fruitful meeting, the main goal was common understanding about the data model and to align with the results from the previous ESSnet.

It was a fruitful meeting, the main goal was common understanding about the data model and to align with the results from the previous ESSnet and the Reference Methodological Framework proposed by Eurostat.

WPI thinks that parallel tracks can be carried out in the hope to find convergence of dealing with the data. WPI takes into account the common technology and methodology proposed by Eurostat.

Furthermore, WPI allocated the different tasks (who is going to do what) and agreed on some deadlines.

Alex praises the work done within WPI on the framework. It also benefits the work in WPK.

*WPI: Marek Cierpial-Wolan*

The first focus was on getting access to data and sort out what kind of data could be available (specifying the big data sources and information about the data, using a template). The first step has been made and now feedback from all partners is awaited.

The second focus was on the development of a draft proposal for software, to be improved and shared with the partners, and for them to give feedback for further improvement. The software is intended to be reusable throughout the ESS. Marek also aims at making the results comparable, although this will not be easy.

Furthermore, work is done together with Germany on webscraping methods to collect tourism data on the internet, even on a micro scale.

Important is the WebEx meeting next week for making decisions about progress, in particular how to go further with the software and webscraping.

Monica emphasizes the work done in the previous ESSnet. In two WPs there were webscraping activities, including the use of webscraping tools and software. It is relevant to understand the relationships between the different projects and in this project between the tracks. It is very good to compare the solutions, describe them and use them within the same framework. In this case it would be very helpful to inform not only WPF and WPK but WPB and WPC as well for usable feedback.

## **5. Matters to be coordinated with WPF**

One activity is very relevant to this track and this is related to business functions. More specific, the definition of big data processes in relationship with business functions.

WPF has started with looking at the work done in the previous ESSnet project like the existing standards (national, European). Currently WPF is working on a document about general information on the business level and the specifications required, but not yet about the way how to use it. In one month the draft will be ready for circulation. The information can also be used for the pilot track.

A lot of material is already on the wiki but it is good to present and share some specific information in track WebEx meetings also. In mid-April this item will be put on the agenda and the link where you can find the information will be sent beforehand.

## **6. Planning of reviews**

Peter mentions that the schedule is not complete yet. New information is still being received. For now there is no need to go through each review date, the most important thing is that the overview covers what is needed for the review process.

Lilli thanks everyone for the information so she can tackle bottlenecks soon enough. Lilli encourages all to look at the wonderful checklist from Faiz when you write the reports. It will not only improve the review and make it easier for the Review Board but also improves your report.

At this moment the Review Board has just two members (Lilli and Anders). If you have potential names please send them to Lilli and Peter. The Review Board is not looking for an expert data scientist but someone who has high level knowledge of big data and has a good overview. In principle there is no requirement where this person comes from, although remuneration has been foreseen only for ESSnet partners.

## **7. Any remaining issues and closing**

Peter is convinced of the use of this meeting. No time was spent on contract or financial issues in this meeting; this type of meeting really complements the regular CG meetings.

The next CG Pilot Track meeting is on the 17<sup>th</sup> of April and the next regular CG meeting is the 20<sup>th</sup> of March. The schedule of the meetings is in the annex of the minutes.

Mark Cierpial-Wolan asks if the prepared report about webscraping methods should be sent for comments to WPB and WPC. This was what Monica meant at agenda point five. Of course WPJ is fully autonomous concerning using the feedback received, but it could be an enrichment.

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

Date	Time (CET)	Issues	How/Where
2019-03-20	10:30 – 12:00 hr*	Regular CG meeting (incl. resources spent)	WebEx meeting
2019-04-03	10:30 – 12:00 hr*	Track CG Implementation	WebEx meeting
2019-04-17	10:30 – 12:00 hr*	Track CG Pilots	WebEx meeting
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