Statistics Netherlands

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

subject	Final minutes CG meeting 4 (Implementation Track) of ESSnet BD II by WebEx 22 February 2019
	2019-03-04

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 and co-chair)	√	Anders Holmberg (Review Board)	√
Johan Fosen (PL WPG)			

Replacements/Guests:

Fernando Reis (Eurostat)	√	Marco Stocchi (Eurostat)	√
Magdalena Six (WPK)	√		

1. **Opening and agenda**

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

   Alexander Kowarik could not attend the meeting and is replaced by Magdalena Six. Albrecht Wirthmann (Eurostat) could not attend the meeting but Fernando Reis replaces him and is also attending for agenda point 4. Also Marco Stocchi form Eurostat attends the meeting but unfortunately, because of audio problems he could not introduce himself.

   Marc has a few remarks about the state of communication. The general structure is online. For each WP there is a brief introduction 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 bottom for more information about the WP.

Essentially 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).

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. The slides from the kick-off meeting (WPF) will be online at the beginning of next week (25th of February).

Fernando asks about the distinction between the public wiki and the private one; how does confluence fit in? Marc answers that the private part, i.e. confluence, is for financial, private and confidential information. The public wiki can also be used for drafts and work in progress, as long as this material is not considered confidential. This part of the public wiki is labelled as working area. It is up to the WP leaders to decide what has to be shielded from the public and to clearly indicate the status of documents and drafts, in order to avoid confusion and misunderstandings.

The agenda was drafted together with Monica, but we start first with the Review Board because Anders has to leave early. In the minutes we follow the order of the original agenda.

2. **WPF state of affairs and matters to be coordinated**

First Monica mentions that she is very satisfied with the group.

There are two main activities. First the design of business functions and the process for implementation in general so this can be used within the ESSnet. On the national level, off course, some specific business functions are necessary. This work is progressing fast and probably Monica can give a general presentation to the WP leaders next month and ask for feedback.

The second activity is the development of a questionnaire which specifically requires input from implementation WPs. The outcome is to understand and model the specific process of each WP to be carried out for the implementation.

Some questions have already been drafted and the aim is that the questionnaire is ready by the end of February. Please give a name to Monica of one contact person to fill in the questionnaire. This will be the first round of the questionnaire and one year later there will be a next questionnaire to have an update. By collecting information and interact with WPs on the work they are doing, the work of WPF acquires a bottom-up dimension.

Peter remarks that he is pleased with this approach, leading to the identification of reusable processes, IT components etc. (reusable for participating and non-participating countries) and framing what is done by the partners in a way that makes comparison possible. This approach may also be applied to methodology and quality.

3. WPB to WPE: state of affairs

WPB: Tomaž

There are currently two main activities. The draft structure of the methodological framework concerning online job vacancies is almost ready. Some comments have been received and the work is on schedule. Second, work to be done is about setting up and getting access to Cedefop data. The communication with them is ongoing. It is essential to know what data can be expected and when. The types of analysis depend on what data is extracted (aggregated data, complete raw data). A WebEx meeting to have a technical discussion is being set up. It is important to define the business case how to use this data (remotely or extracted to local environment).

Peter considers the position of WPB special in the sense that Cedefop can possibly play a role in statistical processes that is unique in the ESS. This has the attention of the ESS Big Data Steering Group, which discussed this in London last week. The ESS Big Data Task Force will talk about this in October in Luxembourg. For Tomaž it is important to integrate Cedefop data and its system but also to pay attention to the development of country specific services. It is very good if at the end there is a fully working system which also can be used by other organizations. At data governance level there might be some fundamental decision to be made.

Monica remarks the importance of this WP from an architecture perspective. The components designed are also useful for others and for interacting on national level. She will follow this closely and stay in contact with WPB.

WPC: Galya

Together with ONS, Bulgaria has been working on the draft structure of a webscraping policy. In the end this must lead to a document including legal aspects of webscraping. The draft version is already on the working area of the wiki. In the first week of March the first final draft will circulate among other members of the WPC for feedback (and other participants may feel free to give comments also). The document will be coordinated with WPB. Also Eurostat will be asked to contribute and give their opinion.

Furthermore, the first draft of the methodological document is almost ready. It is now in review with the partners and will be uploaded at the beginning of March on the wiki (working area). The content of the document at this stage has been kept general but at the end of the project it is not unlikely that national specifics will be added.

Work on the different use cases is ongoing, aimed at getting more details and defining the indicators. New indicators will be included as well.

Monica asks about the Starter Kit and its timing. The Starter Kit is not operational yet, as experimental results will be obtained first. The activity will start at the moment partners are really starting with webscraping and then they can share the difficulties they face, the solutions they use and all other issues to share with partners. The first Starter Kit is planned to be drafted by April next year.

Galya mentions two bottlenecks:

- WPC needs a contact person from ISTAT. ISTAT is a very important partner because they have thorough knowledge from the ESSnet BD 1, but there have been some organizational changes as ISTAT. Solution: Monica will try to provide a new contact person for WPC as soon as possible.
- Up till now Germany hasn't participated in any WebEx meetings. Galya has been in contact with Markus Zwick about this, but still has no name. Peter remarks that Galya can contact Natalie on this, as she plays a central role for Germany.

WPD: Arko

The first draft of the technical report has been ready some time, at this moment comments are edited and the suggestions of Faiz Alzuhail are also looked at. The approach described is the same for all WP partners and can be used by other partners as well, but the implementation at national level may provide some specifics per country in the end. The submission was planned for the first of March but this may not be feasible, it may now come one or at most two weeks later. To Eurostat this would be acceptable.

WPE: Anke

At the end of January there was a fruitful meeting with EMSA about using intermediate products, instead of just receiving the whole pack of data. Within the WP the usefulness of these semi-products will be looked at.

The contract has already been signed between EMSA and Statistics Netherlands to get and use the raw data for the years 2016 and 2017 on European level. This information is also available for ESSnet partners. On national level data is available for the year 2017 for inland waterways.

There will be a face-to face WP meeting next week in Heerlen to discuss the possibilities and ideas.

During the previous project, WPE made use of the Sandbox. As the data is about to be received in May, Anke is wondering what the state of affairs is concerning BDTI (Big Data Test Infra Structure).

4. **Use of BDTI**

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

During the last weeks there were meetings about BDTI on Commission level. The BDTI stalls the received data in the cloud. To access them this has to be scheduled. One cannot get to the data 24/7, this has to be lounged by the support of BDTI.

Concerning EMSA, the idea was floated to put the data on a hard disk and send it to Eurostat, to be uploaded in the cloud with the support of BDTI. It would be much better, of course, if it were possible for EMSA to process and encrypt the data, and upload the information itself. After the data is uploaded the WP members can reach the data by an agreed schedule.

It is necessary to have the requirements from the WPs for the data, software they want to use etc.

There is a little bit of concern with the CG participants, because they want to be able to get at the data anytime, anywhere. With this in mind, WPE proposes to have a meeting on short notice to discuss what the possibilities are.

Fernando makes clear that you cannot compare BDTI with the Sandbox. BDTI is not just a platform like the Sandbox. There are software packages available and also you can select use cases to try out. The infrastructure can be used for processing and also data sets are available for testing. Further support is also permanently available and they can give advice on what a user may want to do which is not standard. You can consult the BDTI team, not only the support desk.

Marc mentions he has created a page on the wiki for the test infrastructure. If Fernando provides him with the slides he will put them on the wiki as background information.

Monica expresses the importance of the opportunity to use the BDTI, especially if ESS processes or tools are shared, such as the tools developed by the previous ESSnet. In the BDTI you can add solutions which can be used by others.

Fernando asks for software to be shared on the platform, to be used by others. He also needs to know what are the needs and requirements for each country. If there would be one use case with all characteristics more or less worked out, this would be very useful.

Peter will also invite Fernando to the next CG meeting (Pilot Track) coming Wednesday, if he is available.

5. **Matters to be coordinated with WPK**

Magdalena mentions that the literature list has been finished and has been sent out to all WPs. The list is also available and reachable in different ways on the wiki. The structure is known by those of the first project, for others please look at the document. The meta information makes the usability of the list much better.

Magdalena wants to emphasize that at this moment there is a lack of papers about quantitative quality indicators. Please inform WPK if you have any such references.

Quality guidelines will now be set up. As the scope of this WP comprises both methodology and quality, Piet Daas (who led the previous project on this subject) will be asked if he can join the meeting next week. If not, Magdalena or Alex may consult him beforehand.

Lilli remarks that it is great what is being done so far in this WP. This ESSnet compares well with what is done on quality indicators in the US, where the orientation is much more on administrative data.

6. **Planning of reviews**

According to Lilli the template which Peter has set up, is very good workable. Thanks so far for the information about the dates for the drafts. The RB can

now schedule to give comments. The peek at the end of the project was an issue at the previous project but now we are earlier with the planning and we have two months extra at the end of the project for rounding up activities.

When you start the report, please keep the list of Faiz in mind while writing your report. Together with the review planning and the template for deliverables and milestones, Marc will put this on the wiki at an appropriate, central place.

The capacity of the RB is low, since for the time being only Lilly and Anders are members of the RB. The RB has to be extended with some people. If WP leaders have suggestions for suitable names, please give them to Lilli, with Anders and Peter in copy. There are some requirements. Some affinity with big data and having an overview of the field would be useful. As there are not many deliverables that are due before summer, we can afford a few months before expanding the RB.

There is some remuneration for RB members so it stands to reason that new RB members come from ESS organizations (including ONS). We can look outside the ESS, but financing may then be more difficult, although contributions on a voluntary basis remain possible.

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 Implementation Track meeting is on the 3rd of April and the next regular CG meeting is the 20th of March. The schedule of the meetings is in the annex of the minutes.

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

Date	Time (CET)	Issues	How/Where
2019-02-27	10:30 – 12:00 hr*	Track CG Pilots	WebEx meeting
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