



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

To WPC partners ESSnet Big Data II

Cc

from Vera Ivanova

subject: **Minutes 2nd WPC meeting ESSnet BD II by WebEx 15<sup>th</sup> of January 2019**  
2019-01-21

**Participants**

Galya Stateva (PL WPC) - BG	✓	Olav ten Bosch - NL	✓
Johannes Gussenbauer - AT	✓	Jacek Maślankowski - PL	✓
Martin Wood - UK	✓	Jussi Ritola - FI	✓
Aidan Condron - IE	✓	Vera Ivanova – technical assistant - BG	✓
Caterina Viviano - IT		Markus Zwick - DE	

The representatives of IT and DE did not participate in the meeting due to other professional engagements.

**Agenda**

The agenda was approved and Galya proceeded on it. She informed that CSO (Ireland) joins the WPC now also as active partner, however without payment. She asked Aidan for their contribution and he confirmed that they intend to work on task 2 and task 3, but still not detailed specifications in which sub-activities they will concentrate.

The discussions on the items on the agenda were as follows:



## 1. Short report and conclusions from Vienna kick off meeting

Galya has already sent information for the main topics discussed at Vienna kick-off meeting and short minutes about Vienna WPC parallel session to all group before Christmas holiday. But she informed again in short the WPC partners that at Vienna kick-off meeting: first, P.Struijs provided an overview of the results of the ESSnet Big Data I. Later, the main characteristics of the new project ESSnet Big Data II were presented. Each coordinator of the WPs of implementation track presented overview of the respective WPs. Galya already downloaded her presentation and integrated report for the kick-off meeting (prepared by Monica) on the wiki page, WPC section ([https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WPC\\_Meetings](https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WPC_Meetings)). Second, regarding cooperation between WPC and WPF - Process and Architecture, after a short presentation of background concepts, WPF coordinator Monica described the objectives of the WP, meaning the definition of reference architectures necessary to carry out big data production both at national and European levels. The specific input to be provided to the WPF by other WPs of the implementation track was detailed in terms of: Information architecture, Application architecture, Solution architecture and Deployment scenarios. Later, Marc Debusschere (WPA) illustrated some relevant point on communication:

- Web site under construction: [https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/Main\\_Page\\_new](https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/Main_Page_new), but now it is available. Before Christmas Galya sent to all WPC team the detailed explanation about wiki status and content. In the beginning of this year, BG team started to fill some content under WPC space on wiki page. You should have in mind that during the working process each WPC country has to upload regularly the results of its work in the Working Area section on the WPC space at the wiki page. If you have need of help or support, please turn to Marc Debusschere or directly to Galya.
- Big Data for European Statistics conference (BDES 2020): To be held in Poland (city to be decided) in November 2020

Olav commented that he had looked at the wiki, and was of the opinion that it would be useful the Dutch IT webscraping software also to be uploaded. Galya confirmed that it would be a good idea to upload this software to be used by others.

Then, Aidan explained shortly the CSO experience – in the field of measuring the Internet economy, e-commerce, OJV, social media – FB, Twitter, bookings online. Victoria will webscrape whole pages, exploring the likes. Galya asked Aidan to share the CSO webscraping software with the WPC partners. He answered it was a technical question, but probably there would not be problem to upload this software on the wiki page.

## 2. Organisation of work and tasks



**Task 1: Webscraping policy** – Galya thanked Martin and Matt that they shared UK ONS webscraping policy, as well as for the advice received from the ethical commission and legal services in ONS. It is better all that to be taken into account in the process of elaborating our Webscraping policy. The work on this task does not start from zero, but as a basic document will be used Netiquette, which has been already elaborated in WP2 by Matt and Olav, as it will be extended and added in such way that the result will be one common policy for webscraping, valid for scraping of characteristics of the enterprises, OJV and tourist websites. This task will be completed only by UK and BG. Not to lose time now Galya talked with Martin later. But anyway, Galya stated that the first draft of the Policy should be ready to the end of February 2019.

**Task 2: Methodological Framework** - Galya stated that the main aim of implementing this WP is to develop and test corresponding prototypes and methodology. As you already know, suitable techniques and concrete methodologies were developed during the pilot phase of the project and implementation will be based on this achievements. The work will consist of the two main tasks: Development of the Methodological framework and defining indicators on enterprise characteristics ,which can supplement existing ICT statistics as experimental statistics and other statistical outputs together with corresponding prototypes.

The idea is to develop and test functional prototypes, following WPF guidelines for the definition of reference architectures necessary to carry out big data production both at national and European levels. It is expected also that results of work will consist of the implementation requirements of prototypes by each use-cases (according to the templates provided by WPF).

### **Use cases: updated and modified**

As Galya already informed via short minutes, during the parallel discussion session in Vienna, old Use cases (WP2, ESSnet on BD II) were updated and modified by participants from AT, BG, PL and UK. Their proposals are the following:

**Use case 1:** URLs Inventory (for all enterprises: without limitations to the enterprises with more than 10 employees) - **AT, BG, IE, NL, PL declared that they intend to work on the Use-cases 1.**

**Use case 2:** Variables in the ICT usage in enterprise survey (Web sales – Ecommerce; Social Media Presence and Job Advertisement; Does the website have any of the following? (price lists, description of goods and services etc.), ICT Polish survey, C10-2018; Does your enterprise gather any information regarding user behavior - ICT Polish survey, C11-2018; frequency of changes on the website, how active on web site) - **AT, BG, IE, NL, PL declared that they intend to work on the Use-cases 2.**

**Use case 3:** Validation of BR and NACE classification - **AT, BG, FI, PL declared that they intend to work on the Use-cases 3. IE will include in it maybe later.**



**Use case 4: Experimental Language Statistics** – it is a completely new use case. **UK will define the new use-case 4 and will work further on it.**

### **Defining of potential statistical indicators**

The name and contents of the statistical indicators/outputs should be adapted to the updated use-cases. As a starting point there are already the defined indicators in WP2. During the kick-off meeting, all participants united around the idea that a clear description of the importance of the indicators is necessary, as well as of the way it is calculated. Galya suggested that each country responsible for defining of the respective use-case to be responsible also for defining of the respective indicators. Each indicator should be clearly defined and its way of calculation to be clearly shown.

Further Olav, Aidan and Galya clearly pointed out that the statistical indicators should be not only those, which replicate the indicators produced by the ICT survey, but it would be better other new indicators to be also defined, thus the usage of web-scraped data would be expanded.

### **3. Physical meetings planning**

All WPC partners agreed on holding the first physical meeting on 17-18 June 2019 in Sofia, BG.

### **4. Other issue - Communication & Cooperation of WPC with WPK**

Expected outcome with respect to Quality issues:

Magdalena and Alex (AT) will prepare a proposal for the structure of the quality guidelines until mid of February 2019. From March until May they will try to fill as much content as possible into the guidelines and then discuss it and finalize until July this year.

The quality report template with a short tutorial how to apply it will be presented by WPK participants in September 2019. The intensions are to have two “filled-out” Quality reports examples available, as one of them will be for webscraping of enterprise characteristics (WPC Alex and Jacek will do that)

Therefore, at the moment, our deadlines for task 5 is still valid, but some fine-tuning of the subtask related with quality have to be done.

### **Talk of Galya with Martin (after the end of the web-ex meeting)**

At the moment the NSI legal department started research of the results achieved by now regarding the legal and ethical aspects of webscraping, as well as studying of the best practice of the UK in this sphere. We expect their position by the end of January 2019. Martin stated that the netiquette based webscraping policy should use best practices of countries’ local precedents. Galya further developed his idea that the Policy should be a more generic European strategic document for all countries, then each country to adapt it in accordance



with national circumstances. Martin agreed that it would be best practices specific for countries but to extract general policy.

### **Conclusions:**

1. BG and UK: The first draft of the webscraping policy should be ready at the end of February 2019.
2. IT and DE will be asked by email about in which use-cases they will participate.
3. The responsible countries for the detailed description of use-cases are as follows: Use case 1 – BG and NL, Use case 2 – AT and PL, Use case 3 - IT? (it will be specified later with the Italian colleagues), Use case 4 – UK.
4. Galya will send the template for description of use-cases.
5. The deadline for use-cases description is by the end of February 2019. It would be better, if possible by the next web-ex meeting to have the first draft of these descriptions in order to discuss them.
6. Galya will send the template for defining of potential statistical indicators. The deadline will be the same as description of use-cases meaning end of February 2019.
7. BG will prepare draft structure of the Methodological framework document and send to all participants by the next WPC web-ex meeting.
8. By the first F2F meeting it is necessary that a real web-scraping should be done and we should have some preliminary results of the processing and analysis of the web-scraped data. Besides we should have approved structure of the Methodological framework and draft version of functional prototypes.

**The next WPC webex meeting is planned for 12<sup>th</sup> February 2019.**