

to Co-ordination Group ESSnet Big Data
cc
from Peter Struijs

subject Final minutes CG 7 meeting Implementation Track of the ESSnet BD II by WebEx of 3 April 2019
2019-04-18

Participants:

Peter Struijs (Chairman)	√	Marek Morze (PL WPH)	
Martin van Sebille (Secretary)		David Salgado (PL WPI)	
Marc Debusschere (Deputy WPA)	√	Marek Cierpiat-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 Japiec (Chair Review Board)	√
Monica Scannapieco (PL WPF)	√	Anders Holmberg (Review Board)	
Johan Fosen (PL WPG)			

Replacements/Guests:

--	--	--	--

1. Opening and agenda

Peter welcomes everybody to the seventh CG meeting of the ESSnet Big Data II, devoted to the Implementation Track.

Martin cannot attend this meeting, so Peter will take care of the minutes. (He will ask Monica to check the first draft.) Anders and Tomaž have also informed Peter about their absence. Tomaž has briefed him. Anke can only attend the first half hour, but she has also briefed Peter. These notifications and briefings are much appreciated.

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

Monica presents the architecture developed by WPF, which comprises a description of the business functions and the lifecycle of processes in which big data are exploited. This architecture is meant to be a common reference for all WPs. The presentation of the architecture and its artefacts is annexed to these minutes. A more elaborate version (with definition or descriptions of terms) will be put on the wiki. Feedback on the draft architecture is most welcome.

Among other things, Monica points out the following:

- The business layer addresses the “what”. At the highest level are the business functions, to which activities and processes are linked. The ultimate aim is to control the flow.
- The architecture makes explicit use of and links to existing standards, in particular GSBPM, GAMS0, CSDA, EARF, NBDRA (from NIST) and CRISP-DM. Each have been assigned their own colour, as is the case for all parts that are new.
- When dealing with big data, it is often not possible to strictly separate the “design” and “production” phases, which are distinguished in GSBPM. For this reason, “development” and “production” are now applied, but it should be noted that, as a consequence, “production” does not have the same meaning as in GSBPM (hence its colour denoting “new”).
- The architecture also explicitly gives a place to new data sources exploration and to trust management. Note that the “support surveys” in the production box play an internal use role.
- Instead of data “collection” the term “acquisition” is used.
- Metadata and quality management in GSBPM meaning are considered.

The discussion makes clear that the draft architecture is very well received. Several remarks or questions are put forward:

- The existing standards may have different statuses. NBDRA is an American standard, to be used only when this is considered to add value.
- What should be the name for the new architecture? For now the acronym BREAL is being used. Thinking aloud, and considering the functional use of colours, Lilli coins the name “Rainbow Model” as a possibility.
- Alex thinks that the notion of “design” may still be relevant. Monica will discuss the way of incorporating the design step, and in particular output design, in the group of WPF.
- There is a need to link the architecture to the architectural efforts of Fabio Ricciato. This will be done at a later stage, when the information architecture is developed.
- Data representation will in due course also need attention.

3. WPB to WPE: state of affairs

WPB:

Tomaž has reported three points to Peter:

- WPB has worked on access to Cedefop data together with Fernando Reis (Eurostat), Vladimir Kvetan (Cedefop) and CRISP.
- The completed WPF questionnaire was sent to Monica.
- A meeting was held with Fernando, Vladimir and Albrecht in order to prepare the ESS Big Data Task Force presentations on OJV (Paris, 4 and 5 April) and to discuss the potential business case for using the Cedefop system at ESS level.

WPC:

The following points are mentioned:

- WPC is making progress according to the planning.
- The answers to the WPF questionnaire were agreed upon; they will be sent to Monica today.
- Concerning webscraping policy, Galya has asked WPB (Tomaž) and WPJ (Marek) for feedback on the document. In mid-April the document will be sent to Eurostat for a reaction.
- The 2019 physical meeting of the WP will take place in June in Sofia.

WPD:

The following points are mentioned:

- WPD is making progress according to the planning.
- The Review Board has provided comments on the draft of the first deliverable. The improved report will be ready by the end of next week.
- WPD will make use of the architecture developed by WPF and Eurostat. Monica would welcome feedback on it.
- The 2019 physical meeting of the WP will take place in June.

WPE:

Anke has reported the following points to Peter:

- Raw European AIS data from EMSA (year 2016 and 2017) has been received.
- On 11 April there will be a technical seminar on AIS which aims to establish the BDTI environment for AIS. After this the AIS data can be uploaded to the BDTI. This technical seminar will take place at statistics Netherlands.
- Milestone EM1: report of the first face to face meeting will be finished at the end of this week.
- The first WPE deliverable (results of the definition phase) will be sent to the Review Board by the end of April (as planned). This week a draft version will be ready.
- During the first face to face meeting, all questions for EMSA were collected (also on possible intermediate products); these were discussed with Albrecht and Fernando. Albrecht has sent these questions to EMSA. If they cannot answer them by e-mail, a meeting will be scheduled.
- The questionnaire sent by WPF will be answered by the beginning of next week.

4. Matters to be coordinated with WPK

Some feedback has been received concerning the literature template.

The midterm meetings of the Implementation Track and the Pilots Track will take place back-to-back in Vienna on 9 to 12 December 2019. The order of these two meetings remains to be established. If there are no objections from WP leaders, the default option is to start with the Implementation Track

meeting. At this CG meeting no objections were raised. At the forthcoming Pilots Track meeting (17 April) this will also be brought up.

5. Any remaining issues and closing

There are no remaining issues. The next general CG meeting will take place on 22 May and the next Implementation Track CG meeting on 5 June (see overview of meetings below).

Peter thanks all participants for their contributions, and in particular Monica for her presentation, and closes the meeting.

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

Date	Time (CET)	Issues	How/Where
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

Annex: Presentation architecture as developed by WPF



