

# KM4 - Report on the Midterm WPK meeting

- 1 [About the Meeting](#)
  - 1.1 [Participants](#)
- 2 [Quality](#)
  - 2.1 [Quality Reporting Template](#)
  - 2.2 [Quality Guidelines](#)
- 3 [Typification](#)
  - 3.1 [Matrix](#)
  - 3.2 [Evolution Roadmap](#)
- 4 [Methodology](#)
  - 4.1 [Structure](#)
  - 4.2 [Additional Examples to \(maybe\) be discussed in the report](#)
- 5 [Literature Review](#)
- 6 [Alignment with BREAL](#)
  - 6.1 [Quality](#)
  - 6.2 [Typification](#)
  - 6.3 [Methodology](#)
- 7 [AOB](#)

## About the Meeting

The meeting was planned to be held on 11 and 12 March 2020 in Rome on ISTAT premises. Because of the current outbreak of the Corona virus all over the world a face-to-face meeting was not possible. Therefore the meeting was split into four video conferences, each on one of the areas where WPK is active:

1. Quality
2. Typification
3. Methodology
4. Literature review and any other business

In general, the meeting was very productive and outcomes were in line with expectation on the face-to-face meeting.

## Participants

- Jacek Maślankowski (GUS, PL)

- David Salgado (INE, ES)
- Sónia Quaresma (INE, PT)
- Gabriele Ascari, Giovanna Brancato, Loredana Di Consiglio and Tiziana Tuoto (ISTAT, IT)
- Piet Daas (CBS, NL)
- Magdalena Six, Alexander Kowarik (STAT,AT)

# Quality

## Quality Reporting Template

- "K6 - Revised Quality Reporting Template" is currently under review and will be finalized until the end of March.
- In the document, which was sent to the RB, the note "this is especially important for the track meeting" and similar notes, which were meant only for participants of the pilot track meeting in December 2019, were still in documents. This must be removed for the final document.
- The document is currently aligned to the draft version of EHQMR. Now that EHQMR is released, K6 should be checked to be aligned with this version. (TODO: AT)
- Some items of the template can be aligned with business functions from BREAL (TODO: AT)
- The layout might be improved if the whole template is in landscape orientation instead of portrait orientation. This will be checked. (TODO: AT)

## Quality Guidelines

- Not all sub-chapters currently have actual prescriptive guidelines. This is criticized in several comments by the RB, both in the first round and also in the second round. However, the decision already made after the first review round was confirmed that some work is not mature enough to give guidelines. We will introduce a comment on this in the introduction of the guidelines. (TODO: AT)
- Magdalena presented differences and similarities with "Quality Guidelines on Multisource Statistics" (QGMS) produced by an ESSnet called KOMUSO. Afterwards we discussed how we can connect these two documents. A new small section (~half page) will be added to the introduction of the guidelines describing the differences and the motivation for the structure of our guidelines (TODO: first draft AT, Review and extend: IT et al):
  - Why is the structure different? Output vs. process oriented
  - QGMS describe the typical situations for admin data. However, in our setting the data classes are more important.
  - The maturity of the used methodology and the whole approach to use such kind of data is very different.
  - Some data configurations can be mapped to applications in the QGBD "throughput phase 2".

- The option of having a more detailed comparison in the final report on quality by the WP was discussed. This should be formulate so it can be used as base for future projects.
  - Coverage comparison
  - Structure comparison
  - Suggestions how to come to a "new" guideline as combination of QGBD and QGMS
- A table for main source of error by class could potential be added to the guidelines, after having a draft, this will be decided in the next Webex meeting (TODO: AT).
- Finally, we went through open comments from the RB and decided on some smaller adaptions and that some comments from the RB will not be implemented, but we will react with a comment.

# Typification

## Matrix

- "K7 Typification matrix for big data projects" will be finalized and delivered to the RB by the end of March (TODO: PT)
- We discussed the structure of the current document and decided on some changes, so the structure will look like
  - Motivation for the Typification Matrix
  - Development
  - Usage
  - Matrix
  - Conclusions
  - References
  - Annex with filled matrix for mobile phone data
- The conclusions will be extended and the insights gained from the track meeting will be presented more prominently (TODO: PT)
- In the development section a comment on the connection to BREAL will be added. (TODO: PT)
- The WPI example will not be updated for K7

## Evolution Roadmap

- The next document will try to relate challenges to quality aspects and treatment to methodologies.
- WPI example will be updated for this deliverable (TODO: ES)
- A overview of different levels of maturity will be produced (in the document this will be highlighted by using different shades of blue). This level of maturity will be defined per data class.
- Elements of the typification matrix will be mapped to business functions of BREAL

# Methodology

## Structure

Starting point:

1. Overview of BD based official statistics processes

1.1 Lessons learned from Google flu Trends

2 Extracting information from new data sources

2.1 Text mining and NLP

2.1.1 Text

2.1.2 Text mining tasks

2.1.2.1 pre-processing steps

2.1.2.2 Entity extraction and more

2.1.2.3 Sentiment analysis

2.1.2.4 Automatic summarization

2.1.3 ESSnet applications

2.2 Image mining

1. Big Data exploration
2. Dealing with errors in BD
3. Combining sources
4. Estimation using BD

6.1 Concept differences

6.2 Representativity in the context of Big Data

6.3 Bias in Big Data

- It was decided to adjust the structure to be more aligned with the quality guidelines, but also in general with BREAL by using the hour glass model. This means that the Estimation section will be split in two parts.
- Linking will be added as a new subsection.

- Penholders/contributors for some (sub)sections:
  - Dealing with errors (AT)
  - Combining sources (IT)
  - Sampling methods (PL)
  - Estimation/inference (ES)
  - Visualization for Big Data exploration (ES)
  - Representativity (IT)
- An updated draft document will be distributed to the partners (TODO: NL) and discussed at the next Webex meeting.

## **Additional Examples to (maybe) be discussed in the report**

The following projects will be checked if they use mature enough methodology to be good examples for the report:

- Sentiment analysis in Italy (TODO: IT)
- Smart meter for vacancies in South Tyrol (TODO: IT)
- Mobile phone data for outbound tourism (FI and AT) (TODO: AT)

## **Literature Review**

- Jacek reported on the current status of the literature overview. Since we last discussed it a couple of new papers were added.
- An online version with search capabilities is in development and will be ready for the final deliverable of the literature review (TODO: PL).
- In CG meetings WPK leader should keep reminding other WPs to provide input for this task.
- The paper "Improving the accuracy of outbound tourism statistics with mobile phone data" by Stats Finland should be added.

## **Alignment with BREAL**

- Sonia Quaresma presented a short introduction into BREAL, the different architectural layers and business functions together with elements of the application layer which are important to our work in WPK.
- We decided that in future communication, we will not talk about alignment or connection to WPF, but with BREAL

## **Quality**

- There will be an effort to align terminology with BREAL, both in the template and in the guidelines, e.g. Collect + Capture → Acquisition and Recording (TODO: AT)

- The hour-glass model for the structure "automatically" aligns the guidelines with BREAL information architecture.

## **Typification**

- In the development section for the typification matrix the connection with BREAL will be explained.
- Elements of the typification matrix will be mapped to business functions of BREAL

## **Methodology**

- Again the hour-glass model for the structure provides some level of alignment.
- Additionally, business functions and elements of the application layers will be linked to sections in the report.

## **AOB**

The next Webex is planned for the last week of March, a survey on availability will be sent out to all partners.