

WebEx Meeting WPJ 13:30 – 14:45 CET, 04.01.2019.

Participants:

PL: Marek Cierpiak-Wolan, Katarzyna Kapica, Teresa Matuła, Piotr Szlachta, Łukasz Zadorożny

NL: Shirley Ortega-Azurduy

IT: Mascia di Torrice

SK: Boris Frankovič + 3 persons

PT: 2 persons

BG: Valentin Chavdarov +1 person

EL: Christina Pierrakou, Dimitrios Stamou, Maria Laftsidou

DE: absent

The meeting began with the welcome of the Leader of the WPJ “Innovative Tourism Statistics” package - Marek Cierpiak-Wolan (PL). Then people participating in WebEx from partner countries presented themselves and their experience in tourism. During the meeting representatives of the Germany (Hesse) were absent.

The presentation created by the team from Rzeszów started the substantive part of the meeting. Its agenda, as well as duration, labor intensity and individual stages of the grant (including deadlines for submitting reports) were presented by the Leader. The next speaker was Kate Kapica, who discussed the types and sources of administrative data sent by the partner countries, which will be used during further work. The analysis of data sources carried out by Kate concerned each country separately and in case of any doubts participants of the meeting were asked for clarification.

The first questions were directed to the representative of the Netherlands team. During the discussion, Shirley Ortega-Azurduy (NL) stated that she did not have data on the nationality of passengers using airlines. They only have the number of people using this type of transport. Water consumption data is available and it will be possible to obtain it quickly. Municipal waste data may be a problem. Shirley Ortega-Azurduy said she would check the possibility of obtaining them when the person responsible for this information will come back to work. In case of mobile data, a representative of the Netherlands stated that they have data from one mobile phone operator, however, these data are not usable. During the discussion Shirley pointed out the lack of information on what kind of tourism countries want to research - tourism referring to the local community, foreigner’s arrivals or residents' trips abroad. Shirley has asked to add a priority column in the table, because each country can have different ones. It should also be determined

whether the data will have short, medium or long-term availability, as well as which specific data we intend to combine with the data obtained through Web Scrapping.

Representatives of the Bulgarian team (BG) stated that they do not have any "Big Data" data sources. Their work under the grant would be limited only to obtaining information through the use of the Web Scraping method.

The persons from Portugal (PT) confirmed that the data called "Financial transactions" means the same as the data "Card transactions". Unfortunately, the data is not yet available, their availability is currently being negotiated with the administrators.

Boris Frankovič (SK) listed the AirBNB report prepared for the Visegrad Group countries as a data source. Slovaks in the near future are to send the abovementioned report to analyze the information contained in it in terms of their usefulness. Data from energy meters, traffic sensors and mobile telephony are currently unavailable. In order to obtain them, it is necessary to negotiate with their administrators. As noted by the Slovaks, negotiations can be very long, which can be a problem considering the duration of the grant and the work planned in it. The Slovak side also does not guarantee the success of these negotiations.

Christina Pierrakou (EL) stated that the Hellenic Statistical Authority (ELSTAT) is going to negotiate with data administrators on the access to data concerning traffic sensors located on the borders. Future negotiations are also planned with data controllers from mobile telephony. ELSTAT would like to focus their web-scraping activities on obtaining data from digital platforms such as Booking.com and AirBNB.

Moreover, as it was stated the Hellenic Tax authority has established a Register of Short-Term Lease Properties posted on digital platform since August 2018. For every lease of such private property the owners have the obligation to submit a declaration of lease with the following information (number of days of leases, income generated, via which digital platform). However, there is a lack of information for number of beds, number of guests and nationality.

So, ELSTAT proposed as additional fields for Web Scrapping the following:

1. **From Airbnb.com:** type of establishment, country of reviewers, location, number of guest and license or registration number
2. **From Booking.com:** type of establishment, country of reviewers, sleeps (number of beds), room type and license or registration number

Although the Italians (IT) officially do not participate in Task 1, the person representing the partner at the meeting promised to check the availability of administrative data sources in their country. Due to staff shortages (holidays), he is unable to determine what data would be possible to obtain. Data from road sensors were mentioned, as well as data on the number of tickets sold in cultural institutions (eg Museums).

In the third stage of the presentation, Piotr Szlachta (PL) took the floor, which briefly presented the most important information about the web scrapping method used. He presented

two portals that was used to test different approach to web scraping data. The method used for portal booking is simulating user behavior to gather data while the portal hotels.com is using hidden api. Piotr also promised to provide partner countries with scripts in order to get acquainted with the methods used in Poland. Leader (PL) stated that an additional WebEx Meeting will be organized during which only the methodology of obtaining data using Web Scraping will be discussed. Then the representatives from the partner countries took the floor.

Shirley Ortega-Azurduy (NL) spoke first and promised to complete and send a table on the methods used. She initially admitted, however, that they use other programming languages than suggested by Piotr (PL). The Italian representative (IT) stated that her colleagues (2-3 people) were doing Web Scraping of "job vacancies", but they never had to deal with obtaining data from tourism in the described way. The next week, they intend to meet with employees responsible for the IT section to verify their potential and capabilities. Christina Pierrakou (EL) representing ELSTAT informed the group that ELSTAT has a relative experience in web scraping of "job vacancies" using ready for use scraping tools. Moreover, ELSTAT supported the idea of a separate WebEx for web scraping issues, as the experienced IT colleague wasn't present during this WebEx.

In the fourth, summarizing part, the representatives of the Polish side answered the questions asked by the partners. The first question was received from Christina Pierrakou (EL), and it concerned the possibility of receiving an updated table extended by the comments of Shirley Ortega-Azurduy (NL). Kate Kapica (PL) confirmed that all countries will receive the revised table by the end of next week and then they will be asked to refill it. Representative of Italian partners (IT) asked if there should be a list of people responsible for Web Scraping. Kate (PL) answered in the affirmative. The Polish side has asked partner countries to provide a list of people taking part in the grant, as well as their distribution in terms of the scope of activities performed in the package. Comments on the start time of the next WebEx were reported by people from Slovakia who would prefer an earlier date (around 12:30). Dates of the next two WebEx meetings were set to the 1st of February and 1st of March.