



BigSurv20 – November 2020 – Utrecht, The Netherlands – virtual conference

Citizen science for official statistics: dream or reality?

Olav ten Bosch¹, Sofie De Broe¹, Kris Vanherle², Ben Laevens¹

1. *Statistics Netherlands, the Netherlands*

2. *Transport and Mobility, Leuven, Belgium*

In today's datafied society people increasingly start measuring phenomena themselves. Especially those that are of particular interest to them, such as the living conditions in their neighbourhood, the energy performance of their house and their own sport performances are popular topics. In many cases the measuring devices used for such activities automatically upload the data to one or more citizen science portals where it is published - usually anonymously - as open data. This poses the question whether this data could be of use for official statistics and, if so, whether the official statistics community should be more active in citizen science communities.

A definite frontrunner in the citizen science area is the air quality community. Over the past 10 years numerous projects were executed where citizens installed an air quality sensor adding data to some central node which collects and displays the data streams as open data. One particular successful project is the *Luftdaten* project [1]. Starting from a relatively small project in Stuttgart it now comprises measurement stations spread across Europe delivering their data to the central hub. Bodies responsible for official air quality monitoring started to take the citizen science data as an additional input for their models, which proves the value in both the initiative as well as the data.

Another example where citizens have a dominant role in data collection is the use of smart cameras for traffic measurements found in the *Telraam* project [2] in Leuven, Belgium. Citizens attach a low resolution camera to the front window of their homes and the embedded open source software - running on a raspberry pi - calculates hourly traffic counts and speeds of bicycles, cars, trucks and pedestrians passing by. The software processes the images from the camera immediately so that they are never stored which prevents privacy problems. The continuous data streams make it possible to show live figures and typical averages per time unit (e.g. hours, weekday, working day/weekend etc.) and changes of averages over time. These data have shown to be of use for many studies in regional mobility patterns.

Yet another example of citizen generated data can be found in the Australian *PVOutput* portal [3]. Citizens or organisations from all over the world - including the Netherlands - connect their solar panel systems to this data portal so that they can monitor and benchmark the power production of their equipment. Research has shown that these high frequency measurement data can be used to build an advanced model to estimate the solar energy generated per region and per day which is a major improvement to what could be published without this citizen science data.

In this paper we dive into various citizen science initiatives from the viewpoint of official statistics. We reflect on the value of the data, the representativity issues contained and we philosophize about the role a statistical office could play in these communities.

[1] <https://luftdaten.info>

[2] <https://telraam.net>

[3] <https://pvoutput.org>