VEO: how official statistics can help prevent emerging infectious diseases

**Keywords:** infectious diseases, Covid-19, social media, mobility, demography

# Introduction

VEO (Versatile Emerging Infectious Diseases Observatory) is a five year European project (2020-2025) with about 20 research institutes in twelve countries around Europe with the aim to predict, signal and prevent infectious diseases such as Covid-19 and others to come [1]. The project was rewarded under the EU call ‘*Mining big data for early detection of infectious disease threats driven by climate change and other factors*’. The project consortium is led by the Dutch Erasmus Medical Centre and includes research institutes on a variety of subjects. Statistics Netherlands is one of the consortium partners, bringing in knowledge on official statistics and (big) data and its legal and ethical considerations.

A key element of the VEO approach is the concept of ‘one health’, indicating that global health is strongly connected with human behaviour, animal health, climate and environmental issues. Therefore an early warning system for emerging diseases has to watch for signals in a wider context. Apart from biodata it is necessary to study contextual (big) data on climate change, farming, animal health, environmental changes, effects of tourism, agriculture practices, but also economical and demographical trends. This is shown graphically in Figure 1. All the surrounding elements have to be studied on indicators of emerging infectious diseases to be included in the VEO monitoring system. Consequently a large variety of data sources have to be taken into consideration.



Figure 1: One health approach in VEO

# Using new data sources

Statistical offices have many years of experience in data discovery and data processing. New data sources have been added next to survey and administrative data as an extra input to the official statistics process. We have learned that the combination of traditional reliable and stable data sources with faster but also more volatile new data sources can result in new indicators. Numerous experimental products have been developed [2]. To help the VEO project discover potential indicators to detect emerging diseases statistics Netherlands actively searched for some – maybe highly experimental but maybe also high potential – use cases. We mention some of the most promising.

## Symptom based social media analysis

Social media analysis is a topic that is actively explored by many research projects around the world and within Statistics Netherlands [3]. However, the main concerns about social media as a source for official statistics are the representativity of the subpopulation on social media and the difficulty of correctly interpreting messages. Statistics Netherlands actively invests in research to both areas. It is on the interpretation of the messages, where Statistics Netherlands investigated within VEO in extracting symptoms of COVID-19 to measure susceptibility in this social media population. Despite the fact that problems concerning the representativity still needs to be solved, the VEO consortium concluded it is worth investigating the value of such instrument for detecting outbreaks. Statistics Netherlands brings in knowledge and experience from former studies on social media analysis. Compared to these earlier projects a more fine-grained approach would be needed where the occurrence of symptoms in messages are taken as a signal for spreads. A VEO result could be a generic and flexible social media analysis engine that can be applied in multiple countries to detect symptoms of newly developing diseases. Work on this idea has started together with VEO partners such as University of Bologna (UNIBO) and the Polytechnical School of Lausanne (EPFL), the latter having experience with social media analysis [4] and annotations using the Crowdbreaks platform [5].

## Changes in mobility patterns

Mobility is one of the fields where new (big) data sources can be found. Traffic loops are one of the first big data sources integrated into official statistics [6]. Data from mobile Network Operators (MNO) has been used in a privacy respecting way to detect local mobility patterns [7]. Access to public transport data is being explored. In the context of VEO it would be valuable to detect local changes in mobility patterns as early as possible. If multiple households in the same neighbourhood have family members staying home instead of commuting to work, schools or social activities, that might be an indication of a disease spreading. While individuals do not even realize they are part of an evolving epidemic, their joint behaviour could be an important signal. To do this it is necessary to develop a detailed model of regular mobility patterns per type of day and time of year (weekdays weekend, holidays) etc. This is something that fits into the mission of a statistical office anyway in order to provide detailed official statistics on mobility. For VEO goals one could imagine to add a system that detects unexplainable local mobility changes. Ideally such a system would work across multiple countries since epidemics to not respect national borders.

## Sewage samples and demography

One of the measures in the Covid-19 crisis is that many countries intensified and improved the sampling of sewage to detect virus spread. Obviously a positive sample does tell something about the people living in the area connected to this particular sampling station. Statistical offices have detailed regional statistics with many population variables. Combining sewage samples with neighbourhood statistics might be valuable to determine whether an infectious disease affects certain subpopulations more than others. Obviously the more detailed the sewage samples are the better the results of such analyses could be. This idea be even more valuable if such analysis could be combined with other measurement data on virus spreads.

## Virus spreading in relation to social network

In an earlier project Statistics Netherlands derived a social network of persons and their relationships from administrative sources [8]. The network contains family relationships, work relationships, people on the same school and people living in the same neighbourhoods. It could be informative to examine how this social network model relates to an infectious disease spreading in reality on an aggregated scale. This can be done by comparing the social network with the actual spreading of earlier outbreaks such as the flue or Covid-19. If it correlates it could be used in future outbreaks to better manage the new epidemic and the measures to be taken locally.

## ELSI

Statistical offices traditionally have good knowledge on ethical, legal and societal implications (ELSI) on the use of survey, administrative data for official statistics. In the last decade this knowledge has been broadened with ELSI aspects of new (big) data sources. VEO partners are specialized in such aspects for medical and biodata. Since the VEO project will use many types of data, linking them together where applicable, Statistics Netherlands assists the consortium on this topic.

# Conclusions

Statistics Netherlands participates in the VEO project targeted at identifying and detecting emerging infectious diseases using biodata and (big) data. Knowledge and experience on big data projects and experimental indicators is applied in the VEO project. Work has been started on the design of a flexible symptom driven social media analysis tool, together with partners actively exploring social media analysis projects. Another idea is to detect changes in local mobility patterns to identify a possible emerging outbreak. A third idea is to combine sewage samples with detailed demographic characteristics to more precisely identify outbreak characteristics. Fourth, it could be informative to use the social network of Dutch population, derived in earlier projects, to study the relation between disease spreading and relationships. In addition to these ideas Statistics Netherlands assists on ELSI aspects. The ideas presented here are in an early stage but we hope they could add to our joint instruments to prevent a next Covid-19-like outbreak.

# References

[1] [https://www.veo-europe.eu](https://www.veo-europe.eu/)

[2] <https://www.cbs.nl/en-gb/about-us/innovation>

[3] Daas, P.J.H., Puts, M.J.H. (2014) Social Media Sentiment and Consumer Confidence. European Central Bank Statistics Paper Series No. 5, Frankfurt, Germany,<https://www.ecb.europa.eu/pub/pdf/scpsps/ecbsp5.pdf> (accessed April 2015).

[4] Salathé et al., Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control, PLos Comput Biol. 2011

[5] <https://www.crowdbreaks.org>

[6] Puts, M.J.H., P.J.H. Daas, M. Tennekes and C. de Blois (2019). Using huge amounts of road sensor data for official statistics. AIMS Mathematics 4: 12–25

[7] Deliverables ESSnet BD I (WP5) and II (WP I), <https://ec.europa.eu/eurostat/cros/content/essnetbigdata_en>

[8] van der Laan, D.J., de Jonge, E.: Measuring segregation using a network of the Dutch population. Paper presented at The 5th International Conference on Computational Social Science (2019)

VEO is a research and innovation action funded by the European Commission’s Horizon2020 programme. Grant Agreement ID 874735.