Remote Sensing Data for Better Statistics

Natalie Rosenski and Clara Schartner

Paper prepared for the 16th Conference of IAOS
OECD Headquarters, Paris, France, 19-21 September 2018

Session 3 A, Day 2, 20/09, 10:30-12:30: New technologies for compiling statistics

Natalie Rosenski
natalie.rosenski@destatis.de
Federal Statistical Office of Germany

Clara Schartner
clara.schartner@destatis.de
Federal Statistical Office of Germany

Remote Sensing Data for Better Statistics

DRAFT VERSION 31/08/2018

Prepared for the 16th Conference of the
International Association of Official Statisticians (IAOS)
OECD Headquarters, Paris, France, 19-21 September 2018

ABSTRACT

New digital data sources and especially remote sensing data change the opportunity to describe social and economic phenomenas. First of all, the digital data sources allow new insides and amend empirical studies based on traditional data. But in particular, the combination of survey, administrative and new digital data sources allows us to get a more comprehensive picture of the society and the economy. Statistical offices as well as research institutes started a lot of common projects in this area. One project that has started in 2015 and will be finished in 2018 examined the use of Copernicus products, that means satellite data, for statistics regarding land cover and land use. However, in this paper the focus is on three on-going projects which combine remote sensing data with traditional data sources to explore social and economic indicators in order to improve official statistics.

The Horizon 2020 project 'MAKing Sustainable development and WELL-being frameworks work for policy analysis' (MAKSWELL) proposes to extend and harmonize indicators able to capture the main characteristics of the 'Beyond GDP' approach proposing a new framework that includes them in the evaluation of the public policies. In particular, satellite data will be used to develop social indicators. Furthermore, satellite information might be used to expand existing small area estimations.

Among other things, the Eurostat project 'Smart Statistics' researches how indicators of the business cycles could be developed from satellite data. Satellite data can be used as auxiliary data to report on economic activity, which is visible from space. This can specifically be the change of activities over time like number of ships and containers at harbours or thermal images of industrial facilities.

MAKSWELL as well as the project Smart Statistics are planning to use data of the Sentinel-2 satellites of the ESA program 'Copernicus'. Both Sentinel-2 satellites are not only free to use, but also provide a global coverage of the Earth's land surface every five days.

The Eurostat project 'Deep Solaris' aims at locating solar panels by developing a machine learning algorithm using high resolution aerial and satellite images. The results of the algorithm will mainly be validated by a comparison to administrative data of solar panels. It is planned to use open source aerial images of a very high resolution in comparison to satellite images of different resolutions.

The paper will give an overview and first results of these projects that started in the beginning of 2018.

Keywords: Satellite data, ESS, Makswell, Deep Solaris, Smart Statistics, Official Statistics

INTRODUCTION

In order to further enhance official statistics especially regarding timeliness, accuracy, relevance and response burden, the exploration of the use of new digital data (also known as big data) for official statistics is essential. For this reason, the Federal Statistical Office of Germay (Destatis) examines different data sources, such as mobile phone and remote sensing data and different techniques such as web scraping. The goal of all these studies is to get a more comprehensive picture of the society and the economy, ideally through the combination of survey, administrative and new digital data.

As Destatis believes that remote sensing data have a huge potential for the production of official statistics, there are several international projects ongoing that explore the use of remote sensing data, especially of satellite data, regarding its use for the determination of different indicators through the detection of different objects.

A first pilot project in this regard is the project 'Cop4Stat_2015plus', which has started in 2015 and will be finished in 2018 and is a cooperation project between the agricultural department of Destatis and the Federal Agency for Cartography and Geodesy (see Wiatr et al. 2016). The project examines the use of satellite data, especially of Copernicus products, for official statistics of land cover and land use.

The focus of this paper, however, is on three on-going projects of the Institute for Research and Development in Federal Statistics of Destatis, which combine remote sensing data with traditional data sources to determine social and economic indicators.

MAKSWELL

The goal of the project 'MAKing Sustainable development and WELL-being frameworks work for policy analysis' (MAKSWELL)[1], is to strengthen the use of well-being and sustainability indicators for policy-making in the EU. One motivation for the realization of this project, which is funded by the European Union's Horizon 2020 Programme, is the still growing attention to 'Beyond GDP' (gross domestic product) indicators, which are proposed to be extended and harmonised in order to use them for public policies. Therefore well-being and sustainability indicators shall be explored if they can be improved by using new digital data sources, such as remote sensing data, and new techniques, such as small area estimation. The project started in November 2017 and will be finished by May 2020. The project leader is the National Statistical Institute (NSI) of Italy (Istat). Also involved are the NSIs of the Netherlands (CBS) and of Hungary (HCSO) as well as the Universities of Trier, Southampton and Pisa. The project contains 8 Work Packages (WPs):

- WP 1 is about the analysis of the frameworks on well-being and sustainability at national and international level for policy making.
- WP 2 deals with the extension of the set of information available on well-being and sustainability by using new data sources in order to be able to derive local indicators.

[1] www.makswell.eu [accessed 31/08/2018]

- WP 3 works on regional poverty measurement as a prototype for modern indicator methodology.
- WP 4 provides time series and multivariate methodology including nowcasting to be applied to well-being indicators and Sustainable Development Goals (SDGs).
- WP 5 focuses on a pilot study for integrated frameworks at different territorial levels and measurements for policy making.
- WP 6 handles the dissemination and communication activities of the project's results.
- WP 7 relates to scientific coordination and project management.
- WP 8 ensures compliance with the 'ethics requirements' set out in this work package.

WP 1 was finished at the end of May 2018. Currently the WPs 2, 3 and 4 are in process.

The first deliverable of WP 1 (for detailed information see Tinto, A. et al. 2018) gives an overview of different initiatives and activities regarding well-being and sustainability at international, national and local level within the European Statistical System (ESS). Although the different activities and initiatives share a common interest in the topics sustainability and well-being, they are not necessarily harmonised in a way that allows comparisons. The compilation of the existing initiatives and frameworks in this matter should answer the questions, which frameworks exist, how they refer to official statistical production and what their use is for policy making in the EU.

In a first step, the dedicated international websites on well-being and sustainability have been consulted. In a second step, the NSIs that are involved in the project have reviewed the 28 members of the European Union regarding their national activities in the area of well-being and SDGs that are used for policy purposes.

The globally most important framework for sustainability is the Agenda 2030 of the United Nations, which was adopted in 2015. It is used by many member states, such as Germany, as a framework for their national Sustainable Development Strategy.

On the European level, there was the EU Sustainable Development Strategy (EU SDS)[2], which has set overall objectives and concrete actions for seven key priority challenges for the period until 2010. Every two years, from 2005 until 2015, Eurostat has produced monitoring reports on the achieved progress of the European Union regarding the targets that were set in the EU SDS[3].

In 2010, the Europe 2020 strategy, which promotes smart, sustainable and inclusive growth in the EU, was adopted by European leaders. The Europe 2020 strategy includes five areas, for which certain targets have to be reached by 2020: 1. Employment, 2. Research & Development and Innovation, 3. Climate Change and Energy, 4. Education, 5. Poverty and Social Exclusion. Eurostat has been involved in the process of defining and monitoring the indicators to support the Europe 2020 strategy.

The EU SDS has been replaced by the EU SDGs. The indicator set is structured along the 17 global SDGs, but is also based on existing indicators like those of the Europe 2020 strategy, the previous EU SDS and the 10 European Commission priorities for 2015-2019[4]. The EU SDG indicator set is limited to 6 indicators per goal and contains 100 indicators in total. It is adjusted to the UN list of SDGs, but does not cover all aspects of the global SDGs.

---

[2] http://ec.europa.eu/environment/sustainable-development/strategy/index_en.htm [accessed 31/08/2018]
[3] https://ec.europa.eu/eurostat/web/sdi/eu-sds [accessed 31/08/2018]
[4] https://ec.europa.eu/commission/priorities_en [accessed 31/08/2018]

As a custodian agency of some SDG indicators, the OECD is responsible for the monitoring of these SDGs and also involved in developing frameworks to measure SDGs. But foremost, the OECD is leading the global researches on well-being with the OECD framework for Measuring Well-being and Progress[5].

Research on well-being is based on the understanding that purely economic indicators, such as the gross domestic product (GDP), are not sufficient to explain people's quality of life. The Report by 'The Commission on the Measurement of Economic Performance and Social Progress' (also known as Stiglitz-Sen-Fitoussi Report)[6] and the 'Beyond GDP' initiative of the European Commission[7] are important milestones in this matter and have set the ground for further actions such as the researches of the OECD.

Regarding the review of the 28 members of the European Union with respect to their national activities in the area of well-being and SDGs, so-called 'country profiles' were compiled.

According to the answers of the NSIs, 19 out of the 28 EU countries have developed a national well-being framework and 27 have developed a national SDG framework. 11 of the national well-being frameworks are used in the national policy cycle and in 12 countries well-being indicators are available below the national level. Regarding SDG indicators, in 21 member states, they are used in the national policy cycle and in 12 countries SDG indicators are available below national level.

The goal of the second deliverable of WP 1 was a comparison of the databases on 'Beyond GDP' initiatives within official statistics or rather the frameworks, which were developed in 19 of 28 EU countries to measure well-being on a national level. The comparison to the OECD 'How's life?' initiative showed that the topics proposed by the OECD were covered in almost all countries.

The next step in the project MAKSWELL includes an overview of the status of SDG indicators in the member states Italy, Netherland and Germany, which are involved in WP 2. By information such as if the indicator is still under development, the current method and data sources used to determine the indicator and possible alternatives, it will be possible to identify the indicators that are most suitable to be improved by new digital data.

Together with Destatis' German partners the Federal Agency of Cartography and Geodesy (BKG) and the German Aerospace Center (DLR), Destatis will be able to explore the use of satellite data for the improvement of SDGs in the European Union, such as the UN SDG indicator 11.1.1 'Proportion of urban population living in slums, informal settlements or inadequate housing'. In this regard, urban poverty and people living in inadequate housing could be examined through measuring the spatial structure of these settlements and the combination with official statistics. The approach of combining remote sensing data with other data sources to identify areas of urban poor has already been studied for developing countries (see e.g. Taubenböck et al. 2018 and Friesen et al. 2018). While this methodology still has to be adapted for EU countries, it can serve as a first framework for further examination.

[5] http://www.oecd.org/statistics/measuring-well-being-and-progress.htm [accessed 31/08/2018]
[6] https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report [accessed 31/08/2018]
[7] http://ec.europa.eu/environment/beyond_gdp/index_en.html [accessed 31/08/2018]

SMART STATISTICS

The Eurostat project ‘Smart Statistics’ is a project that Destatis conducts in cooperation with the service provider Sogeti Luxemburg and the Slovenian Institute ‘Jožef Stefan’ (JSI). The project started in February 2018 and will be finished in March 2019. At the end of January 2019, Eurostat will organise a workshop on ‘Trusted Smart Statistics: policymaking in the age of the IoT’[8]. The project includes three Proof of Concepts (PoCs). The JSI will explore ‘Smart Mobility Statistics’ (PoC1) and ‘Smart Labour Market Statistics’ (PoC3). Destatis is specifically interested in ‘Smart Business Cycle Statistics’ (PoC2) and will explore how economic indicators can be derived from satellite imagery.

Business cycles are important economic phenomena. The Gross Domestic Product (GDP) for developed countries occurs in cycles around a positive trend. These cycles have an enormous influence on society’s welfare and well-being. Basically, business cycles are the workload of the economic production factors labour and capital. The workload of the production factor labour is highly correlated to the employment rate and with this to the income of most of the households. Fluctuations in using the factor capital have an influence on the investments or the income of the capital owners for example. All these effects are able to reinforce or to stabilize business cycles and with this to influence the growth of the GDP.

Because of the influence of business cycles for income and wealth, the economic parameters that are responsible for the cycles are of core interest to politicians. Their goal is to stabilize the growth of the GDP through economic policy. For this purpose they need information about the state of the business cycle. This can be done for example by forming indicators of the business cycle and combining them into a system (see van Ruth et al. 2005). Furthermore, it is important that this information is of high quality and up-to-date.

Traditional methods of reporting the GDP in official statistics work very well and have a high accuracy. However, the reporting process is complex and introduces a time lag of several weeks to publication. The goal of ‘Smart Statistics’ is to reduce this time lag by deriving indicators from economic activities, which are visible in satellite images. Satellite images are available with a short delay of only a few hours. The processing of the data and the detection of economic activities can also be done comparatively fast and thus allows a publication of economic indicators with a delay of only a few days. However, while these indicators are based on auxiliary data and cannot be expected to have the same accuracy as traditional methods of determining the GDP, these indicators can help to determine the state of the business cycle in almost real time.

The economic activities need to be detectable from space, which means that they have to take place outdoors or leave traces outside. Containerships for example are clearly visible. The revenue of supermarkets does not actually take place outdoors, but the number of parked cars gives an indication and is traceable with very high resolution satellite images. Several enterprises such as ‘SpaceKnow’[9], ‘Orbital Insight’[10] and ‘RS Metrics’[11] estimate the revenue of retailers by analyzing the number of cars parked in

[8] https://ec.europa.eu/eurostat/cros/content/workshop-trusted-smart-statistics-policymaking-age-iot_en [accessed 31/08/2018]
[9] https://www.spaceknow.com/ [accessed 31/08/2018]
[10] https://orbitalinsight.com/ [accessed 31/08/2018]
[11] https://rsmetrics.com/ [accessed 31/08/2018]

front of their stores in combination with socioeconomic factors. By this, they are able to provide almost real time estimations to the stores.

The indicators that are investigated in the PoC are the following:

- Number of ships and containers in harbours
- Number of planes at airports
- Filling levels of fuel tanks
- Number of cars in front of supermarkets
- Number of ships on rivers and canals
- Length of traffic jams during rush hour
- Construction sites

One disadvantage of using optical satellite imagery are clouds. Objects cannot be detected with clouds or shadows of clouds. On average 55% of the land is covered by clouds with seasonal and spatial variation (King et al. 2013).This means that all indicators are heavily affected by clouds, but the extent depends on the season. An alternative to optical remote sensing imagery are radar sensors which are independent of the weather, but can only detect certain objects such as ships.

The objects, which need to be detected, are of different sizes and thus the required minimal resolution varies. The freely available Sentinel-2 images with a spatial resolution of 10 m do not allow to detect some of the necessary objects, such as cars (see Abraham, L and Sasikumar, M. 2014).

There are commercial satellites, which can provide a higher spatial resolution. At the moment, imagery with a very high resolution such as 1 m is very expensive. The exact price depends on whether the image has to be ordered , which means that the image would otherwise not have been taken, and how up-to-date the images are. Since the most recent images are of interest to form near real time indicators, this data is very expensive. Satellite images for several areas are necessary to derive enough information from different places to construct a conclusive indicator for a country or region, which makes the process even more costly.

Furthermore Sentinel-2 has a temporal resolution of 5 days which is a problem when there are clouds in subsequent images which would mean that no observations can be reported in that time period. It would be possible to combine the images of different satellites to achieve a higher temporal coverage, however this would raise the cost and workload significantly.

While the calculation of indicators does not require the development of new methods, there is only limited scientific research on constructing economic indicators with remote sensing data. There are, however, several private enterprises, which are successfully using information derived from satellite images to detect the magnitude or change in economic activity. Unfortunately, their research is commercially valuable so that no details of their methods are publicly available.

The main obstacle at the moment is the high cost of very high resolution images and its infrequent coverage of areas. However, several private enterprises, such as BlackSky[12] and Axelspace[13], plan on

[12] https://www.blacksky.com/ [accessed 31/08/2018]
[13] https://www.axelspace.com/en/ [accessed 31/08/2018]

launching multiple small satellites, which can be used in combination to get a much better temporal resolution. With more affordable prices for high resolution, both in spatial and temporal resolution, the economic avitivies needed to construct these indicators will be more feasible to detect. Therefore, the smart business cycle statistics should become more accurate, affordable and reliable.

DEEP SOLARIS

The Eurostat project ‘Merging statistics and geospatial information in Member States (Deep Solaris)’ started in January 2018, in cooperation with CBS, the NSI of Belgium (Statbel), the statistical office of North Rine-Westphalia (IT.NRW) and the Dutch Institute ‘Business Intelligence & Smart Services (BISS)’. The project will be finished in January 2020.The goal is the (semi-) automated analysis of satellite and aerial images for energy transition and sustainable indicators.

In a first step, the institute BISS is developing and training an algorithm by using machine learning techniques to detect solar panels in aerial pictures and satellite imagery. The second step is the validation of the results, generated by the algorithm, via administrative data. The final step is the production of a map of solar panels using the algorithm. Furthermore, the potential of harmonizing the registry for solar panels across the EU will be discussed.

Within the project, all the work is done for small test areas in the Netherlands, Belgium and Germany. The overall goal is to develop an application that could be used by the other member states of the European Statistical System to detect solar panels.

Currently, a first report about the available datasets is drafted by the project partners.

Until now, only aerial images have been used by BISS to train the algorithm, but using satellite data will be the next step. In the Netherlands and Belgium, aerial images with a 25 cm resolution are updated once a year and provided via open data web services. However, the large coverage which is needed for this project is not feasible since, at least in the Netherlands, there seems to be a limit of 10,000 requests per month so that CBS uses their own datasets for this project. In Germany, aerial images with a resolution of even 10 cm are available for free, but only for a couple of federal states, e.g. for North Rine-Westphalia.

In order to test the possibility to detect solar panels by satellite data and in order to explore which the required minimum resolution is for the solar panel detection, it was decided in the project to use satellite images with the following resolutions: below 1 metre, 1 to 2 metres, 5 and 10 metres resolution. Satellite data with a very high and high resolution, of below 1 metre up to 5 metres, are commercial and have to be purchased. Only satellite data of 10 metres resolution is freely available. The data with 10 metres resolution is provided by Sentinel-2, which is a satellite of the Copernicus programme. Copernicus is the European Union's Earth Observation Programme, that offers information services based on satellite Earth Observation and in situ (non-space) data[14]. Sentinel-2 consists actually of two satellites through which the whole surface of the earth can be scanned every 5 days. Furthermore, in comparison to commercial satellites, images do not have to be ordered to be taken, but are constantly taken and stored. However, it is

[14] http://copernicus.eu/main/overview [accessed 31/08/2018]

expected that data from Sentinel-2 can only be used to identify large solar panels. For this reason, a first commercial data set with a resolution of 50 cm is about to be purchased and used in this project.

CONCLUSION

Besides the exploration of the use of remote sensing data for agricultural purposes, Destatis currently engages in three different projects that examine the use of remote sensing data for official statistics. The project ‘Makswell’ is about the improvement of well-being and sustainability indicators that are useful for policy-making. The PoC 3 of the project ‘Smart Statistics’ that Destatis explores is investigating ‘Smart Business Cycle Statistics’. The project ‘Deep Solaris’ is about using remote sensing data to develop an algorithm, which is able to detect solar panels.

The algorithm of the project ‘Deep Solaris’ is trained for a specific goal, which is the detection of solar panels, but the lessons learned from this project are very helpful for multiple purposes. First of all, it can be used as an example process of developing an algorithm with remote sensing data for the detection of objects. Furthermore, different resolutions will be tested for detecting the same object which provides guidance for future decisicions on the correct resolution. This choice is of key importance, since it determines the cost of the object detection: If the resolution is chosen too high, the satellite imagery is unnessecarily expensive, while a resolution that is too low renders the image unuseable.

The project ‘Smart Statistics’ will also set the ground for further projects and analysis on the use of remoting sensing data for official statistics. Remote sensing data seems to have a lot of potential, e.g. for improving the time lag of reporting the GDP in official statistics. However, remote sensing data that is able to detect objects, e.g. ships in harbours, is quite expensive currently and only available on special request. Remote sensing data that are available for free and have a high temporal resolution are probably not able to detect objects due to their spatial resolution.

The conclusions of the project ‘Smart Statistics’ and ‘Deep Solaris’ will also be helpful for the project ‘Makswell’ when specific indicators will be explored regarding their potential to be improved by remote sensing data.

Even with some constraints regarding the access to and costs of remote sensing data that seems necessary to improve official statistics, these data have a lot of potential for different statistics because of their actuality, accuracy and relevance.

BIBLIOGRAPHY

Abraham, Lizy, Sasikumar, M. (2014): Vehicle Detection and classification from high resolution satellite images. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume II-1, ISPRS Technical Commission I Symposium, 17 – 20 November 2014, Denver, Colorado, USA.

Friesen, John; Taubenböck, Hannes; Wurm, Michael; Pelz, Peter F. (2018): The similar size of slums. In: *Habitat International*, *73*, pp. 79-88.

King, Michael D., Platnick, Steven, Menzel, W. Paul, Ackerman, Steven A., Hubanks, Paul A. (2013): Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites. In: *IEEE Transactions on Geoscience and Remote Sensing* 51 (7), pp. 3826–3852.

Ruth, Floris van; Schouten, Barry; Wekker, Roberto (2005): The Statistics Netherlands' Business Cycle Tracer. Methodological aspects; concept, cycle computation and indicator selection. In: *Statistics Netherland report*.

Taubenböck, Hannes, Kraff, Nicolas J., Wurm, Michael (2018): The morphology of the Arrival City - A global categorization based on literature surveys and remotely sensed data. In: *Applied Geography* 92, pp. 150–167.

Tinto, Alessandra; Bacchini, Fabio; Baldazzi, Barbara; Ferruzza, Angela; Rondinella, Tommaso; van den Brakel, Jan A., Willems, R.M.A.; Rosenski, Natalie; Zimmermann , Thomas; Andrási, Zsolt; Farkas, Máté; Fábián, Zsófia; Münnich, Ralf; Ertz, Florian (2018): International and national experiences and main insights for policy use of well-being and sustainability frameworks. *A collective work within the EU funded project MAKSWELL for the 16th conference of IAOS.*

Wiatr, Thomas; Suresh, Gopika; Gehrke, Ralf; Hovenbitzer, Michael (2016): Copernicus – Practice of daily life in a national mapping agency? In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B1, 2016 XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic.