



Economic and Social Council

Distr.: General
17 April 2019

Original: English

Economic Commission for Europe

Conference of European Statisticians

67th plenary session

Paris, 26-28 June 2019

Item 2 (a) of the provisional agenda

New data sources – accessibility and use

Session 1: Accessing new data sources

Access to Big Data for statistical purposes

Note by the Federal Statistical Office of Germany

Summary

The document provides an overview of the activities of the Federal Statistical Office of Germany (Destatis) to ensure access to new digital data sources. Many challenges still need to be solved to use big data. First, there are quality and methodological issues, such as coverage, selection bias as well as the timely and regular availability. Destatis has already started several feasibility studies in this area. To be able to use new digital data for official statistics on a more sustainable basis, a permanent access to data held by private enterprises is necessary. Therefore, Destatis is aiming for a legal basis to get access to privately held data to be able to build further cooperation.

This document is presented to the 2019 Conference of European Statisticians seminar on “New data sources – accessibility and use” for discussion.



I. Introduction

1. The Federal Statistical Office of Germany (Destatis) as well as the working group “New digital data in official statistics” of Destatis’ statistical advisory board is convinced that big data in combination with survey and administrative data have a huge potential for official statistics. It is expected that official statistics will be more timely and more precise with these new data sources and that they will cause a reduction of response burden, especially for private enterprises, whose data is collected through surveys so far. By this, official statistics meet societies’ expectation of a digital era that is to use all information sources available and to avoid resourceful surveys if possible. If official statistics can save money by this approach, is still to see. Possible cost reductions in survey areas might be overcompensated – at least short-term – by investments for hard- and software as well as the training of statistical employees. Also, even if the access to privately held data is guaranteed legally, the pre-processing of this data for statistical purposes will maybe (still) require NSIs to pay for this service. Specific cost-benefit analysis still needs to be done, before big data is used for the production of statistics and not only for feasibility studies.

2. To examine the suitability of big data for official statistics, Destatis is involved in several feasibility studies, e.g. the exploration of mobile phone and satellite data for various purposes regarding official statistics. Regarding the accessibility and processing of big data, Destatis is in cooperation with different organisations, such as the mobile phone company T-Systems International GmbH and Motionlogic GmbH (which are both subsidiaries of Deutsche Telekom AG), the German Aerospace Center and the German Federal Agency for Cartography and Geodesy. Furthermore, Destatis is a member of the European Statistical System (ESS) Task Force, the ESS Steering Group as well as the United Nations Global Working Group on Big Data for Official Statistics and cooperates with a lot of other National Statistical Institutes and Universities in regard of different European projects.

3. But in order to use big data, many challenges still need to be solved. First, there are quality and methodological issues, such as coverage, selection bias as well as the timely and regular availability. Then the internal and external communication regarding the use of big data has to be intensified. The required IT infrastructure and the required interfaces for the integration of semi-final products of private data owners have to be examined as well as the required skills for using big data. Furthermore, there are different legal questions that need to be solved with different stakeholders such as ministries, data protection officers at national and regional levels, the Federal Network Agency and last but not least the data providers.

4. The main focus of this document is the access to big data, which is crucial for official statistics to produce better and more precise statistics with less burden for respondents and which is one of the major obstacles as experiences from big data projects have shown.

II. Approach of the statistical offices

5. From Destatis’ point of view, data of private enterprises, which generally arise as a by-product of their key business, should be accessible for public purposes that means for the purposes of official statistics. Therefore, a permanent access to data held by private enterprises is necessary, which is unattached of their willing or the market situation. In this regard, Destatis is aiming for a legal basis to get access to privately held data, on which further cooperation can be premised on.

6. Regarding a legal regulation, Destatis and the State Office for Information and Technology of North Rhine-Westphalia (IT.NRW) commissioned a comparative law report about the access to big data for public purposes in Europe. Thereby, it can be guaranteed that German official statistical agencies are producing statistics under the same conditions as the partners in the ESS. Besides, the report is supposed to examine how the legal framework of German federal statistics has to be changed to be able to use big data in the statistical offices in Germany.

7. On the European level¹, in 2016 Eurostat assigned a contract to analyse the legal situation concerning access, use and dissemination of statistics derived from big data for all the member states of the European Union. So far, there are two benchmarks within the EU regarding access to privately held data. The first is France with its “loi numérique”², which obliges companies – under certain conditions – to open up data that is defined as “public interest data”, e.g. for the French National Statistical Institute (INSEE). The second is the Digital Economy Act 2017 of the United Kingdom³ (see part 5, chapter 7, section 80, new section 45D) that will enable the UK’s Office for National Statistics (ONS) to access data held by large and medium-sized businesses and charities for one or more of their statutory functions, i.e. the production of official statistics and statistical research.

8. Besides, in 2016 the Global Working Group on Big Data for Official Statistics published “Recommendations for Access to Data from Private Organisations for Official Statistics”⁴ which can be used as guidelines for negotiations with private enterprises which have big data sources that are valuable for official statistics.

9. Furthermore, in April 2018 the European Commission published a Communication “Towards a common European data Space”⁵ and a guidance document on sharing private sector data in the European data economy⁶, which is a part of their Digital Single Market strategy and refers to data sharing between businesses and the public sector. The European Commission supports the idea of improving access to privately held data by public authorities such as National Statistical Institutes, but also keeps the interest of the private enterprises in mind, such as compensation for the companies’ investments for adapting data.

10. In October 2018, the European Statistical System Committee (ESSC) adopted the Bucharest Memorandum and has recognised “the need to further develop the legal framework at European and national level to reduce obstacles to the access, use and integration of heterogeneous data for sustainably producing Trusted Smart Statistics.”

11. Generally, the European General Data Protection Regulation (GDPR) is supposed to increase trust in digital services and their security by defining common standards for data protection and by securing the fundamental right regarding informational self-determination. The GDPR is applicable since 25 May 2018 in all member states to harmonise data privacy laws across Europe.

12. A partially processing of data by private enterprises could also simplify the compliance with data protection legislation. Statistical intermediate products on an aggregated level, without person-related information, are mostly enough for official statistics and will make it easier for data protection officers to accept the use of big data for official statistics. In order to use statistical products that are pre-processed by private enterprises, clear and binding standards have to be developed. On the European level this process has already started and is called “Trusted Smart Statistics”. In this regard, statistical

¹ See unpublished version “Legal initiatives for data access” from Eurostat, in the context of the Task Force Big Data, April 2017.

² See on the Legifrance site: <https://www.republique-numerique.fr/pages/in-english> and <https://www.republique-numerique.fr/pages/digital-republic-bill-rationale> (both accessed: 18 May 2018).

³ See on the UK’s National Archives site: <http://www.legislation.gov.uk/ukpga/2017/30/contents/enacted> and http://www.legislation.gov.uk/ukpga/2017/30/pdfs/ukpgaen_20170030_en.pdf (both accessed: 18 May 2018).

⁴ See on the UN Statistical Division site: <https://unstats.un.org/unsd/bigdata/conferences/2016/gwg/Item%20%20%28i%29%20a%20-%20Recommendations%20for%20access%20to%20data%20from%20private%20organizations%20for%20official%20statistics%20Draft%2014%20July%202016.pdf> (accessed: 18 May 2018).

⁵ See on the European Commission site: <https://ec.europa.eu/digital-single-market/en/news/communication-towards-common-european-data-space> (accessed: 18 May 2018).

⁶ See on the European Commission site: <https://ec.europa.eu/digital-single-market/en/news/staff-working-document-guidance-sharing-private-sector-data-european-data-economy> (accessed: 18 May 2018).

intermediate products of third parties have to fulfill qualitative, ethical and methodological standards so that they can be combined with survey and administrative data and be used for the production of official statistics.

III. Possible solution in Germany

13. From today's perspective, big data itself should continue to be free of costs for official statistics, regardless of the kind of data. That means that big data should be provided from respondents without a reimbursement of costs.

14. A legislation in this regard could be similar to the existing legislation concerning the use of administrative data. For the assessment of the suitability, the administrative bodies have to transmit information about the origin, structure, content and other metadata to the statistical offices if requested. If a Federal Ministry then decides that further studies should be conducted, the administrative bodies also have to transmit anonymised, individual data (that is without names and addresses) to Destatis. If the suitability of the data is determined, there needs to be a specific legal regulation in the law that mandates the particular statistic.

15. Furthermore, it still needs to be decided if the statistical offices of Germany will make use of services of the private enterprises for data processing and will reimburse the enterprises for these services. The fact that data processing by private enterprises will partly be more cost-efficient, indicates that cost sharing by the statistical offices or a cooperation between the private enterprises and the statistical offices is probably the best option. In case of a cooperation between private enterprises and the statistical offices, the principle of neutrality of official statistics has to be assured and to be secured through legally binding, transparent procedures.

IV. Additional (legal) questions

16. In addition to the access to big data, there are other legal questions that still need to be discussed, e.g. if there is a specific data protection legislation needed for data that private enterprises will provide to official statistics. In this regard, it still needs to be explored if the findings or products that arise from cooperation with private enterprises could lead to an unfair competition and how sensitive business information can be protected.

17. Other questions relate to the ownership of big data e.g., which is especially difficult to answer when it comes to data that are generated by machines, e.g. in the agricultural sector. Also to be tackled are ethical questions, such as if the use of big data affects the professional standards of official statistics.

18. The full potential of big data will probably only be apparent when they are combined with survey and administrative data. Therefore, lawmakers should also lay the foundations for the necessary legal requirements regarding the combination of different data sources, especially regarding person-related data.

19. The data of official statistics are currently made available for the science by the Research Data Centres of the Federal Statistical Office and the statistical offices of the Länder, in consideration of the legal requirements of section 16 subsection 2 of the Act on Statistics for Federal Purposes. The data availability for science has to be possible for big data as well in the future. Also, the results of official statistics have to be open to scientific evaluation.

V. Conclusion

20. A permanent access to big data is crucial for the production of official statistics on a regular basis. This must not be affected by current market situations and the willing of private enterprises for sharing their data.

21. To maintain the sustainability of official statistics, Destatis is convinced that the legal framework of German federal statistics needs to be revised concerning the access to privately held data especially in times of general digitisation.
