

CONFERENCE OF EUROPEAN STATISTICIANS

Approved

Meeting of the 2020/2021 Bureau
Geneva (virtual), 15-16 October 2020

Item IV (f) of the Provisional
Agenda

FOLLOW-UP TO THE 2019 CONFERENCE OF EUROPEAN STATISTICIANS
SEMINAR ON ACCESS AND USE OF NEW DATA SOURCES

Note by the Secretariat

*The note presents the action plan developed by a Task Team as a follow-up to the CES 2019 seminar 'New data sources – accessibility and use', and a summary of ongoing international work in these areas. The Bureau approved the action plan in October 2019 and asked the Secretariat to follow-up. **The Bureau reviewed the ongoing activities related to the action plan and considered the areas to be well covered. Some additional suggestions are included in the comments.***

CONTENTS

I.	INTRODUCTION.....	1
II.	ACTION PLAN AND ACTIVITIES UNDERTAKEN	2
	A. Communicating the importance of data access to policy makers.....	2
	B. Case studies of sustainable partnerships with the private sector	3
	C. Technical standards, including a focus on geographic detail	4
	D. Identifying the jobs, skills and strategies to advance the use of new and alternative data	6
	E. Identifying cross-country and cross-organizational pilot projects related to skills needed for new and alternative data	7
	F. Identifying best practices to address legal hurdles to alternative data sources.....	8

I. INTRODUCTION

1. The June 2019 Conference of European Statisticians (CES) seminar on access and use of new data sources recognized the opportunities and challenges that statistical offices have encountered in using new data sources, and the skills needed to use these data sources successfully. The Conference proposed to develop an action plan focusing on a set of key issues related to data access and data skills.

2. A Task Team¹, led by Switzerland and the United States developed an action plan identifying a series of possible activities to take the work forward. The Bureau discussed the

¹Denmark, Slovenia, Sweden, Switzerland, US, the High-Level Group on Modernization of Official Statistics (HLG-MOS), the United Nations Statistical Division (UNSD), and Prof. Frauke Kreuter (University of Maryland and University of Mannheim).

plan in October 2019 and noted that it provides a very useful structure for grouping actions related to new data sources and for setting priorities. For example, Eurostat has already used this structure to give an overview of their activities undertaken in this area. The Bureau also noted that it would be useful to make the results of work on accessing new data sources available in one place, as the topic continues to be discussed in many international fora.

3. The Bureau asked the Secretariat to contact the groups mentioned under the action items, clarify whether they can take up the proposed tasks, and report to the CES Bureau. This should provide a basis for the Bureau to coordinate the implementation of the action plan, discuss concrete follow-up actions and their schedule.

4. Meanwhile, the Covid-19 pandemic has in some cases opened up the use of additional data sources, including privately held data. The question is how to maintain this access after the pandemic is over, and to capitalize on the opportunities presented during the Covid-19 crisis. The CES 2020 plenary session considered some of these issues in relation to data stewardship and noted that ‘There should be a balance between the use of private data sources and retaining trust in authoritative official statistics. Legislative changes and public debate may be needed to clarify the use of privately held data for official statistics.’

5. **The note presents the work areas and activities from the action plan and information about ongoing international work in these areas. This information is presented in italics after each action.** The text in blue provides web links to more information.

II. ACTION PLAN AND ACTIVITIES UNDERTAKEN

A. Communicating the importance of data access to policy makers

6. The political, legal and public opinion context differs by country and topic, and changes continuously. The most practical way forward is to learn from the experience of other countries and to build on the outcome of the HLG-MOS Strategic Communication Framework Project.

Action 1: Development of a communication concept with examples of important statistical information, which could be made available to decision-makers:

- (a) Collect case studies and share examples on (UNECE can identify the best platform to share);
- (b) Add case studies and examples to the Strategic Communication Framework on communication aspects of using new data sources. HLG-MOS is already gathering such artefacts;
- (c) Countries can take action and use the HLG-MOS Strategic Communication Framework, and prepare or adjust their communication strategies to stakeholders (policy makers and the general public) that include effective crisis and issue management plans for using new data sources.

Updates and comments:

Canada: the topic of the ethics, opportunities and challenges of accessing and using new data sources – which could extend to enhanced guidance for the Fundamental Principles of Official Statistics in that regard - should be addressed.

Eurostat: Communication to policy makers and the public may be coordinated with regional and local stakeholders in particular with respect to using privately held data for official statistics of a finer granularity such as regional.

Moreover, alliances with other governmental stakeholders would be beneficial hence communication to policy makers and the public can be coordinated in the context of “data driven public administration” and “national data policy”. In this context, the community of statisticians in official statistics should leverage the reputation of statistical offices as regards quality and trustworthiness and protect statistical independence.

OECD: In collaboration with the HLG-MOS BSTN group, it has been decided to organise a workshop on this topic in the context of the HLG-MOS series of webinars. The workshop, named ‘StatsBot – A chatbot for official statistics’, will showcase the results of a proof of concept carried out jointly by OECD, CBS and StatCan with start-up. It will also share on the research work by INRIA research agency on related topic, guided by OECD. The workshop will be an opportunity to draw on these experiences so as to envisage the possibility of joint investment in data accessibility and user research.

HLG-MOS: An on-line version of the Strategic Communication Framework has been created. The HLG-MOS Capability and Communication group will update and extend the framework and welcomes countries to submit case studies that can be shared with the statistical community and decision-makers. CES members are encouraged to provide case studies.

B. Case studies of sustainable partnerships with the private sector

7. While around the world statistical authorities in the public sector are still mostly regarded as the primary source for data and indicators, the private sector is increasingly moving into the business of providing data analyses and projecting trends. Evidently, the private sector is moving much faster than the public sector in using available Big Data sources. Partnerships with the private sector in the use of Big Data for official statistics seem to be one of the most attractive ways forward.

Action 2: Organizations are encouraged to add and share case studies to the Strategic Partnerships in Official Statistics wiki on best practices in establishing partnerships with alternative data providers. Work should be consistent with the UN Global Platform which is establishing partnerships with commercial data and service providers. Countries should further consult work done in this area:

- (a) Modern partnership framework for official statistics ([ECE/CES/BUR/2018/OCT/8](#));
- (b) [Guidelines for the establishment and use of partnerships in Big Data Projects for Official Statistics](#);
- (c) [The Strategic Communication Framework Phase Two](#) that covers communication with stakeholders
- (d) The outcome and follow-up activities of Strategic Partnerships in Official Statistics, the [in-depth review on strategic partnerships by the CES Bureau](#), and the outcome of the [High-level seminar on Strategic Partnerships in Official Statistics \(11-12 April 2018\)](#)

- (e) The outcome of the CES Bureau request for feedback and ideas for possible follow-up to the report on strategic partnerships ([ECE/CES/2019/42](#))
- (f) Best practices in establishing partnerships with alternative data providers
- (g) Develop a transnational approach to cooperation with multinational companies

Updates and comments:

Additional resources (suggested by OECD):

- (h) [Public-Private Partnership: Lessons learned, future steps \(Paris21, 2016\)](#)
- (i) [WB/IMF initiative on development data partnership](#) – after initial steps in assessing the development data partnership, OECD is now considering joining it. OECD believes the WB/IMF framework provides a strong basis for global engagement with GAFAs and digital platforms in the area of non-profit data sharing agreements.

New Zealand: New Zealand can provide two case studies: (1) on partnerships with the private sector drawing on the Data Ventures experience prior to and during the covid-19 pandemic; and (2) sharing of lesson learnt from our experience as an NSO meeting new demands for information during the covid-19 pandemic. UNECE will include the case studies provided on their virtual Strategic Communication Framework platform.

Mexico: It would be useful to take advantage of new data sources in combination with the conditions that COVID-19 poses new challenges and make an inventory of emerging initiatives as a result of COVID-19.

Poland: There are possible synergies between the work undertaken by the UNECE and the work of the European Statistical System within the Partnership Group Task Force on access to privately-held data, or other bodies dealing with the topic within different international statistical systems.

HLG-MOS is continuing the work on the Strategic Communication Framework. It does not have the capacity to take on the other topics. CES countries are requested to make further resource available and to submit activity proposals to the HLG-MOS for considerations for their work programme.

This action will also be addressed by the first [UNECE Expert Meeting on Modernizing Statistical Legislation](#) (3 and 5 November 2020), described under section F. 'Identifying best practices to address legal hurdles to alternative data sources' (p.7). The meeting will include a discussion on partnerships with private data providers during Covid-19 pandemic and what is needed for such partnerships to continue while returning to "normal".

C. Technical standards, including a focus on geographic detail

8. The widespread adoption of Big Data and the Internet of Things (IoT) has increased manifold over the last few years. Industries, technologists and home users recognise the importance of Big Data and IoT in their lives. However, the rapid growth of Big Data and IoT may raise some concerns, such as security, authentication and access control.

9. In this context, it is needed to consider what the consequences will be of adopting an “edge computing” paradigm to the use of IoT for official statistics. In other words, what quality assurance mechanisms should be adopted for bringing the data processing operation to the point of data collection (through sensors or other means).

Action 3: Create an inventory of technical standards for using Big Data. For example, prior work in this area has identified such standards as:

- (a) Data compiled using alternative sources meet the same quality standards as data currently published by national statistical offices;
- (b) Alternative data are reliable and consistent;
- (c) Alternative data are correctly linked to or blended with survey data;
- (d) The use of alternative data sources is documented in a transparent manner;
- (e) The IT infrastructure can support the collection and use of alternative data.

10. Big Data and its use in official statistics is covered by numerous groups, including: 1) the UN Global Working Group on Big Data and 2) the ESSnet Big Data (Eurostat) project. To avoid duplication, organisations should consider the work done within these frameworks when they propose new Big Data activities. With respect to machine learning, organizations can draw on the HLG-MOS Machine Learning project to address issues and apply machine learning guidance.

Updates and comments:

Eurostat: Quality assurance mechanisms should be preceded by the development of new methodological frameworks (hourglass model) for using specific big data classes in official statistics. For example, the use of Mobile Network Operators signalling data, collected primarily for network operation processes (lower part of the hourglass model involving communication experts), that can be repurposed to infer statistics on patterns of human mobility (upper part of the hourglass model involving statisticians).

OECD is working towards developing a renewed data quality framework, taking into account new data sources. Preliminary version could be shared by the end of Q2 2021.

HLG-MOS: The experiences and outputs of the HLG-MOS Machine Learning project are shared on the public [Machine Learning for Official Statistics](#) digital platform. In collaboration with ONS, the project will continue as an interactive community platform to continue collaboration and the sharing of experiences. Activities will be coordinated with the United Nations Global Working Group on Big Data and the ESSnet Big Data project. CES members are encouraged to actively join the community platform.

11. The demand for more diversified, sophisticated and rapid statistical services could be met by leveraging the emerging sources of Big Data, such as those relating to remote sensing imagery, transactional and social media data and mobile device data. Statistical agencies around the world have a strong interest in investigating the viability of using satellite imagery data to improve official statistics on a wide range of topics spanning agriculture, the environment, business activity and transport. Satellite imagery has significant potential to provide more timely statistical outputs, to reduce the frequency of surveys, to reduce respondent burden and other costs, and to provide data at a more disaggregated level for informed decision making.

Action 4: How the use of alternative data can yield statistics with greater geographic detail (for example, data that incorporate postal codes or GPS coordinates). The 2020 CES seminar on *New roles for statistical and geospatial agencies in emerging national data ecosystems* benefitted from the discussion with heads of national mapping and cartographic agencies due to the joint session with the UN Committee of Experts on Global Geospatial Information Management (UN-GGIM: Europe).

Updates:

CES 2020: The joint CES – UN-GGIM: Europe session ‘New roles for statistical and geospatial agencies in emerging national data ecosystems’ explored the implications of national data ecosystems for national statistical and geospatial agencies. The session concluded that geospatial and statistical agencies can support each other to enhance their roles, the dialogue between the two communities should continue, and joint activities with concrete outcomes should be undertaken. The joint session recognized the need for future work in several areas, such as determining where geospatial and statistical standards need to be better aligned to ensure greater interoperability, and coordinating capacity development activities. It was recommended to hold another joint plenary session in June 2022.

OECD: OECD is carrying out several activities overseen by the OECD Working Party on Territorial Indicators on the use of unconventional data for new and more geographically granular evidence in a number of domains. A workshop was organised on this topic in November 2018 with publicly available proceedings: <http://www.oecd.org/cfe/regional-policy/modernising-statistical-systems.htm>. Among the relevant examples are the Canadian experience of mapping the building footprint through partnership with the private sector, using information derived from satellite imagery. Another example is the delineation of labour market areas using mobile phone data to infer the daily mobility of people.

12. The [UN Global Working Group on Big Data for Official Statistics](#) has a dedicated [Task Team on Satellite Imagery and Geo-Spatial Data](#), which aims to provide strategic vision, direction and development of a global work plan on utilising satellite imagery and geo-spatial data for official statistics and indicators for Sustainable Development Goals. The Task Team is building on precedents to innovatively solve the many challenges facing the use of satellite imagery and geo-spatial data sources, and could be mandated to identify how the use of alternative data can yield statistics with greater geographic detail (for example, data that incorporate postal codes or GPS coordinates).

D. Identifying the jobs, skills and strategies to advance the use of new and alternative data

13. Work in this area is being done by the [Training, Competencies and Capacity Development Task Team](#) of the UN Global Working Group on Big Data for Official Statistics. The HLG-MOS Group on Developing Organisational Capabilities has done work on Competency Profiles for Big Data Teams, Guidelines on Risk Management Practices in Statistical Organisations, and various related products (refer to <https://statswiki.unece.org/x/UgAMDQ>) and is currently working on a Training Framework. Again, we encourage collaboration across groups to avoid duplication of effort.

Action 5: The groups can be requested to see if the following topics are covered and if not, be included in their work:

- (a) Inventory of changes in organizational structures to support incorporation of new data sources into the statistical process
- (b) Develop a common job description for data scientists
- (c) Develop innovative personnel policies to retain highly trained data analysts
- (d) How countries are bringing together different disciplines to solve problems – IT, mathematics, statistics, economics

Updates and comments:

Eurostat: The methodological framework of individual (big) data classes for developing official statistics explicitly requires bringing together resources from different disciplines. Therefore, the specific profiles should be identified within the respective methodological frameworks. E.g., the methodological framework for using Mobile Network Operator data for developing spatial-temporal statistics on human presence and mobility would require experts in communications, mobile networks, GIS and statistics.

HLG-MOS: The HLG-MOS Capabilities and Communication group lacks resources. CES members are encouraged to join the group and work on these topics.

E. Identifying cross-country and cross-organizational pilot projects related to skills needed for new and alternative data

14. The HLG-MOS Group on Developing Organisational Capabilities can be requested to see if the following topics are covered and if not, be included in their work.

Action 6: Identify training required for staff to acquire appropriate skills to work with new data sources (for example, build a data science curriculum). The UN Global Working Group on Big Data for Official Statistics and the ESSnet Big Data (Eurostat) project have various pilot studies. In the area of machine learning, the HLG-MOS Machine Learning project has various pilot projects. Organisations can join these efforts to:

- (a) Identify joint pilot projects using alternative data sources
- (b) Review case studies of new data sources addressing old issues

15. Through such identified projects and case studies, required training can be identified.

16. Recommendations to be considered in the future, based on their relevance and other ongoing efforts.

Updates:

HLG-MOS: Nineteen pilot studies have been conducted under the HLG-MOS Machine Learning project in Classification and Coding, Edit and Imputation and on Imagery. Summary results of these studies as well as quality aspects and strategies to integrate Machine Learning in the production of official statistics are shared on the [Machine Learning for Official Statistics digital platform](#). The UK Data Science campus may provide support to continue the HLG-MOS machine learning community.

F. Identifying best practices to address legal hurdles to alternative data sources

17. Big Data is by definition different from traditional data sources currently used by national statistical systems. Having access to new data sources is of great concern but linked with a lot of legal hurdles. Another consideration is the necessity and proportionality aspects and how they are critical in ensuring public acceptance of privacy protection.

Action 7: Establishing best practices regarding how legal hurdles related to alternative data sources have been addressed and how new laws and regulations are helping with access to alternative data should be done. This may also require assessing (1) the role of NSOs in national data stewardship, and (2) the role of privacy preserving techniques to alleviate privacy concerns.

Updates:

CES (June 2020) considered data stewardship at its 2020 plenary session. The meeting emphasized the need to continue the discussion, sharing ideas and experiences. It was proposed to set up a task force to prepare a roadmap and action plan to support NSOs.

HLG-MOS: The HLG-MOS Input Privacy-Preservation project and the Synthetic Data Sets Group under the HLG-MOS [Blue Skies Thinking Network](#) work on making data available in line with confidentiality requirements.

UNECE: The first [UNECE Expert Meeting on Modernizing Statistical Legislation](#) (3 and 5 Nov. 2020) will explore the relation between the right to privacy and the right to live in an informed society, and the legal aspect of data access, use and exchange, including accessing new data sources during the Covid-19 pandemic. The meeting will bring together statisticians, representatives of data protection agencies, academia and private sector to discuss how to create a culture in which people and organizations are willing to share private and public data with NSOs and what safeguards need to be in the legislative framework and beyond so that people are reassured that their data entrusted to NSOs are protected. The meeting will take place virtually on 3 and 5 November.

* * * * *