**Supplementing the United Nations
Fundamental Principles of Official Statistics: Mapping and Guidance for the United
Nations Fundamental Principles of Official Statistics against non-conventional and non-
traditional data sources**

Prepared by: Friends of the Chair group on the Fundamental Principles of Official Statistics

Supplementing the United Nations
Fundamental Principles of Official Statistics: Mapping and Guidance[1] for the United Nations Fundamental Principles of Official Statistics against non-conventional and non-traditional data sources

A. Background

In modern statistical systems, a lawful right of access to administrative and secondary data sources cannot be neglected. Without that right, national statistical systems would not be able to meet the growing demands of users.[2] Going further, an important element for an efficient functioning of national statistical systems are infrastructures that make it possible to share data and to interlink them, particularly those stemming from different sources and stakeholders.

Background for this situation can be found in today's reality, the so-called "data revolution". Indeed, the data industry has been expanding very fast with data providers proliferating, thereby challenging the traditional data producers, National Statistical Office (NSOs), to redefine themselves and their data production systems. In addition, the value of pace in getting the information has been outgrowing the value of quality thereof.

Users frequently turn to sources other than official statistics, unaware of the quality limitations of what they are receiving. Policy makers, businesses, citizens want not only to be informed, they want to be informed quickly and in an easily accessible way. While having progressively acquired and adopted new technologies, methods and standards, NSOs cannot remain solely providers of good-quality official statistics. Their roles are evolving towards the ones of story-tellers, data stewards, data integrators and finally, quality assessors and providers of standards and secure data architecture, especially in the era of fake news and post-truths.[3]

Public statistics are currently facing new needs resulting from phenomena changing previously known social and economic structures. Advancing globalization and digitization is forcing official statisticians to cover new areas, change existing surveys and finally explore the potential hidden in new data sources. Moreover, the data revolution is also impacting social attitudes to data sharing, privacy and confidentiality. Official statistics has been striving to regulate access to private sector data for statistical purposes[4], however the social perception may differ in this case from the attitude taken to statistical

---

[1] This background document highlights guidance and recommendations for the integration of non-conventional and non-traditional data sources into national statistical systems, while adhering to the Fundamental Principles of Official Statistics. It elaborates on Part II of the background document *Supplementing the United Nations Fundamental Principles of Official Statistics* presented at the 50th Statistical Commission and available at: https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3b-FPOS-Implementation-guidelines-E.pdf.

[2] See: Statistical Capacity Building for Sustainable Development: "While national statistical legislation should define and legislate for all aspects of a statistical system, there are three absolutely necessary components: the professional independence of that system and in particular of the head of that system; The NSO and other statistical agencies of the NSS must be legally empowered to collect primary information for statistical purposes from persons, households, businesses and other institutions and have legal access to all appropriate administrative microdata held by national and regional public administrations for the purposes of compiling aggregate statistical information; and the safeguarding of confidentiality must be clearly set out in law". In: *Developing the fundamental pillars necessary for modern national statistical systems*/Statistical Journal of the IAOS 33 (2017) 895–909 895 DOI 10.3233/SJI-160331 IOS Pres.

[3] Anil Arora, opening speech from the High-Level Workshop on Strategic Partnerships, Geneva, 11 April 2018.

[4] To this end, a B2G Data Sharing Expert Group has been created at the European level as an informal body whose main task is to assist the European Commission in assessing issues connected with business-to-government data sharing. It is managed by Commissions' Directorate-General for Communication Networks, Content and Technology (DG CNECT) The governing principles for B2G data sharing were outlined in 2018 by the Commission in its Communication 'Towards a common European data space'. Those are accompanied by details legal and practical considerations set out in Guidance on private sector data sharing.

access for public sector administrative data. Therefore, the existence of the reference principles describing universal key values of the official statistics is of utmost importance.

The second global and vital challenge of today is the 2030 Agenda for Sustainable Development (2030 Agenda) where NSOs fulfil the task of supporting data needs, as "[n]o country is at the point of being able to report on 100 percent of the indicators, not even the most sophisticated among us. The 2030 Agenda has brought a sense of humility and camaraderie between national statistical offices. It is an opportunity to bring producers and consumers together around an important cause and demonstrate the critical role of NSOs and the relevance of our science and craft".[5]

The trends outlined above indicate the necessity of introducing innovative methods and techniques to produce official statistics, using and combining many (new) data sources, including those called non-traditional or non-conventional. The non-traditional sources cover big data (mobile phone data, social media, etc.), citizen generated data (CGD), and a range of other so-called new data sources. In each of their cases, we deal with external data handlers, and in the case of big data, most often with private sector companies.

As coordinators of the National Statistical Systems (NSSs), NSOs need to improve their own capacity in adopting technologically advanced ways of including the new data in their production processes, which requires meeting several conditions. These are related to obtaining the consent of the data owners for their use, while complying with applicable law (e.g. protection of personal data), durability of the data source, data compatibility with the official statistical production process or its change and adaptation, ensuring methodological regime and data quality. In addition, NSOs are faced with the challenge of developing modern systems that incorporate standards, quality and interoperability in the data production processes.

Statistical surveys and administrative registers are and will remain important sources to produce official statistics, as it is confirmed, among many others, by examples from the United Kingdom or New Zealand[17]. Yet, the enhancement and enrichment of the traditional data gathering methods by big data analytics appears inevitable.

Bearing this in mind, opening up privately held data of general interest for re-use by national statistical systems is undoubtedly beneficial for society. However, the lack of a uniform legal basis determining conditions under which privately held data can be made available for official statistics hinder the success

---

[5] Anil Arora, opening speech from the High-Level Workshop on Strategic Partnerships, Geneva, 11 April 2018 17 The United Kingdom uses a range of sources to produce official statistics. While the key sources are surveys and administrative datasets, the UK Office for National Statistics (ONS) has been conducting some preliminary experiments with new data sources including web scraped price information, geo-located Twitter traces, smart energy meter data and aerial photography. The summing up of the mentioned works is available at:
https://www.ons.gov.uk/aboutus/whatwedo/programmesandprojects/theonsbigdataproject More detail is given below:
a)   The use of online property databases to classify housing types, particularly caravan sites, with a view to improving our on-the-ground intelligence for census enumeration. See
https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/ons methodologyworkingpaperseriesno11identifyingcaravanhomesinzoopladatajune2017
b)   Web scraping price data from retailers with a view to generating more timely pricing information to feed into price indices.
c)   Preliminary trials with anonymous, aggregated mobile telephone data to see whether we can estimate commuting flows
d)   The ONS Data Science Campus is doing wider research around how big data can serve the public good – see https://datasciencecampus.ons.gov.uk/projects/
New Zealand uses a range of sources to produce official statistics, most are based on government administrative data sources and surveys. However, data from credit card providers, data from electronic payments switch companies (all electronic card transactions), scanner data for consumer electronics and web-scraped prices are also being used. A trial to use utilities data has been done, being however an ad hoc example to understand population movements during the Canterbury earthquakes. Statistics New Zealand is continuing to explore a range of different sources to support the production of official statistics.

of this venture. The arguments advanced by data owners against granting access to their data for statistical offices concern mainly property rights, confidentiality and trade secrets.

In view of the above, the need to reconsider the meaning of the UNFPOS resurfaces. Official statistics must meet the test of practical utility, regarding its indispensable character for the information system of a democratic society. Statisticians need to follow strictly ethics in the collection, processing, storage and presentation of statistical data, along with the transparency of methods used.[6]

Including non-traditional and non-conventional data sources in the statistical production constitutes another challenge for the applicability, timeliness and relevance of the UNFPOS. As it is almost impossible for policy makers to differentiate between official and non-official statistics, a challenge to communicate quality is being faced by NSOs.[7] Stemming directly from the Principles, the professional independence of statistical offices and their longstanding experience in dealing with personal and confidential data is widely appreciated and perceived as providing extra assurance with regard to the further use of the data owned by private companies.

The following mapping of the UNFPOS against the mentioned sources aims at proving that those features remain unchanged. For the purposes of the document, notions "new data sources", "non-conventional data sources", "non-traditional data sources" and "big data" are used interchangeably.

Although, no uniform definition has yet been agreed upon for big data, it is broadly accepted that the notion refers to data sets of increasing volume, velocity and variety; the 3 V's.[8] In the paper "An Assessment of big data for official statistics in the Caribbean", a fourth V is added to describe big data, namely their "veracity", conformity to facts.[21] Another feature of those data sources is their lack of consistent structure, meaning the lack of a pre-defined data model and/or the fact that they do not fit well into conventional relational databases. Of interest for official statistics, this type of data represents considerable challenge about its inclusion into the statistical production due to ethical concerns, law and ownership-related issues.

Following the classification of the High-Level Group for the Modernization of Official Statistics (HLGMOS), established in 2010 under the umbrella of Conference of European Statisticians, the large data sources comprise:

- Administrative data (arising from the administration of a program, be it governmental or not), e.g. electronic medical records, hospital visits, insurance records, bank records, food banks, etc.
- Commercial or transactional sources (arising from the transaction between two entities), e.g. credit card transactions, on-line transactions (including from mobile devices), etc.
- From sensors, e.g. satellite imaging, road sensors, climate sensors, etc.
- From tracking devices, e.g. tracking data from mobile telephones, GPS, etc.
- Behavioural, e.g. online searches (about a product, a service or any other type of information), online page view, etc.
- Opinion, e.g. comments on social media, etc.[9]

---

[6] Dominik Rozkrut, keynote speech from the Conference of European Statistics Stakeholders, Bamberg, 18 October 2018.
[7] Sylvie Michaud, TA2.14 Non-official Data: Challenges and Opportunities for NSOs, Canadian Experience, UN World Data Forum, Dubai 2018.
[8] *What does "big data" mean for official statistics?,* United Nations Economic Commission for Europe, Conference Of European Statisticians, 10 March 2013What does big data mean for official statistics?, UNECE, CES, 10 March 2013 [21] Abdulkadri Abdullahi, Evans Alecia, Ash Tanisha, *An assessment of big data for official statistics in the Caribbean: Challenges and opportunities,* January 2016
[9] *What does "big data" mean for official statistics?,* UNECE, CES, 10 March 2013

Another type of classification, developed by the UNECE Task Team on Big Data[10], provides a division into <u>human-sourced information</u> (social networks, blogs and comments, personal documents, pictures, videos, internet searches, mobile data content, SMS, user-generated maps and e-mail) <u>process-mediated data</u> (data produced by public agencies, medical records, data produced by businesses, commercial transactions, banking/stock records, e-commerce, credit cards) and <u>machine-generated data</u>, so called *internet of things*, i.e. data from sensors, home automation, weather/pollution sensors, traffic sensors, mobile sensors, mobile phone location, security/surveillance videos/images, mobile locations, and satellite images.

Irrespective of the classification adopted, it is evident that the use of new data sources in official statistics represents a set of considerable challenges falling into at least one of the following categories:
- Legislative, i.e., with respect to the access and use of data;
- Privacy, i.e., managing public trust and acceptance of data re-use and its link to other sources;
- Financial, i.e., potential costs of sourcing data versus the benefits;
- Management, e.g., policies and directives about the management and protection of the data;
- Methodological, i.e., data quality and suitability of statistical methods;
- Technological, i.e., issues related to information technology.[11]

Considering the above, UNFPOS emerge as the main reference and a solid basis for considering ethical issues related to the inclusion of the new data sources in the official statistics. In 2017 Europe's *Ethical Guidelines*[12] have been published, aiming to draw the attention of statistical authorities to possible issues of professional ethics that can appear with the use of big data in the production of official statistics. Based on the results of the projects carried out by NSOs and research, investigating different angles of the use of big data in official statistics, the *Guidelines* recommend an approach that is also compliant with the statistical code of conduct. For the purposes of the present mapping exercise, the *Guidelines'* approach of examining the possible issues at three main stages of the statistical production process is applied.

Principle 5 is considered to be the most applicable of the UNFPOS to the new data sources[13].

> *"Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents".*

It sets up the framework for new data sources to be considered as a natural supplementation to statistical production. Provided that non-conventional sources ensure the quality of statistical output are cost-efficient and minimize the reporting burden for the data providers, the tendency is to increase their use in the official statistics. Building on that, the examples of using sensor data, mobile phone data, and satellite images can be given across countries.

At all main stages of the statistical production process—acquisition, processing and dissemination— questions of an ethical nature concerning the cornerstone values of official statistics are naturally raised.

---

[10] https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data
[11] *What does "big data" mean for official statistics?,* UNECE, CES, 10 March 2013
[12] Services concerning ethical, communicational, skills issues and methodological cooperation related to the use of Big Data in European statistics, Task 1. Ethical Review, Deliverable 1.3. Report on ethical guidelines June 2017 https://ec.europa.eu/eurostat/cros/system/files/draft_ethical_guidelines_final.pdf.
[13] See F. Perucci, *Implementing the Fundamental Principles in a transforming statistical system*, 15th IAOS Conference, Abu Dhabi, UAE, 6-8 December 2016

Guidance to National Statistical Offices for data acquisition

While acquiring data, the problem of data ownership emerges. Big data are mostly collected by private companies, which are usually not bound by the law in most countries to make data available for official statistics. Therefore, the provision of data is often based on strategic partnerships established with private companies to gain access to data. Here the main ethical reference should be Principle 1 called the impartiality principle and ensuring the right for information.

> *"(...) official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information".*

This principle stresses the need to avoid any pressure from businesses to put their interest above the public interest. Secondly, the selection of partners should be done with full transparency and be preceded by thorough research. From practical examples, given by the members of the Group, strengthening the Research Divisions was the recurrent one. Another form of ensuring that the use of new data sources is consistent with Principle 1 is a preference given to a multi-mode approach while collecting the information. Validation of information and constant crosschecks have also been given as examples thereof.

Relevance, impartiality, equal access, interaction with users and planning being the most important features covered by this principle, a risk of competition between national statistical offices and other stakeholders to access new data sources has been mentioned as a factor to be avoided.

Therefore, according to the Principle 6—the confidentiality principle—NSOs need to assure big data providers that their data will be used exclusively for the purposes of official statistics and there is minimal risk of harm to their business. The confidentiality principle affords a distinct advantage for national statistical offices compared with other data providers. In order to work consistently to this principle, NSOs should be informed by data providers whether their customers are aware that the data about them can be delivered to statistical authorities.

Other examples of actions to be undertaken in respect of the applicability of UNFPOS to the use of new data sources, at the stage of data collection, are actions of a legislative nature. Within Principle 6, all sorts of adherence to national, international and regional codes have been quoted as examples. In light of new data sources, strict adherence to confidentiality has been described as the most important factor, due to the possibility of inadvertent disclosure of confidential information deriving from a new data source. A case of social media use has been evoked as an example of risk and difficulty in obtaining users' consent to the usage data containing information about them.

Recommendations

1. When entering into strategic partnership arrangements with private sector data suppliers and intermediaries, NSOs need to ensure that they put the public interest ahead of their partners' commercial interests.
2. The strategic partners need to be well-researched and selected based on a thorough and transparent process so as to maintain public trust and confidence.
3. NSOs need to ensure that there is an appropriate legal framework in place and they work with regulators like a Privacy Commissioner to maintain public trust and confidence.
4. NSOs need to ensure that they have the right capabilities and capacity eg research, and engagement to support strategic partnerships for data acquisition given the additional complexities associated with using non-conventional and non-traditional data sources.
5. NSOs need to have a plan and approach to introducing new non-conventional and non-traditional data sources ideally developed with users to ensure their needs are being considered and met.

6. NSOs need to ensure that they act in good faith and use the data exclusively for official statistics purposes and disseminate in a way that minimizes harm to the partner's business.
7. NSOs need to ensure that the customers of private sector suppliers and intermediaries are aware of any data sharing arrangements and the conditions thereof. This includes legal obligations and privacy protections as appropriate.

Guidance to National Statistical Offices for data processing

At the stage of data processing, the UNFPOS appear as a remedy for the major threat, which is the number of quality-related issues that may be compromised while integrating big data with the official statistics. Examples: risk of bias and manipulation within big datasets, no guarantee in stability and continuity of data structure, lack of scientific proof while using statistical models or imputation techniques for the data processing.

The key reason being that big data in general are not designed for statistical purposes and therefore they do not comply with statistical definitions, standards and methods.[14] According to Principle 3 *("To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics"),* it is of utmost importance to use new data sources according to verifiable and internationally comparable, transparent standards and procedures.

Members of the FOC FPOS Group also highlighted the increasing importance of metadata and paradata, as well as the need for establishing and agreeing upon new quality attributes for new data sources, if they differ from the existing ones. This would emphasize the professionalism of statistical offices, which is one of their major assets that makes official statistics a reliable source of information.

During big data processing a risk of revealing personal information is also pertinent, as well as improper use of those data that can damage the reputation of the official statistics. Therefore again, Principle 1 (on relevance, equal access and impartiality), Principle 6 (on confidentiality) and Principle 2 ("*To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.")*, provide for the importance of legislation, standards and ethics to prevent loss of trust - may be widely applicable.

Recommendations

1. NSOs need to have a well-considered and rigorous quality assurance framework for the use of non-conventional and non-traditional data sources use in official statistics which has ideally been developed and tested with users and/or has their needs in mind.
2. NSOs need to provide an appropriate level of technical and process documentation including metadata and paradata to support the current processing and analysis of non-conventional and non-traditional data sources used in official statistics.
3. NSOs need to ensure there are appropriate confidentiality methods in place and that these meet the required standards for official statistics.
4. NSOs need to have risk management and contingency plans in place to address supply issues eg late delivery, and errors that could damage the reputation of official statistics.
5. NSOs need to maintain a good working relationship at the strategic and operational levels with the partners' organisation in order to ensure they are fulfilling the intent of the strategic partnership and actively solving problems and challenges.

---

[14] Report on ethical guidelines, p. 5. https://ec.europa.eu/eurostat/cros/system/files/draft_ethical_guidelines_final.pdf

Guidance to National Statistical Offices for data and statistics dissemination

At the stage of dissemination of statistical information, faced with complex techniques needed for the production of statistical output while using new data sources, it is vital to: inform the users about methods and procedures used to produce statistics (Principle 3), adhere to international standards ensuring quality and comparability (Principle 9 – *"The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels".*), as well as to communicate and educate users, preventing an erroneous interpretation of statistical information (Principle 4 "*The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics").* Based upon that, new data sources should be duly described, and applied methods and models should be documented to support an independent assessment of data processing and statistical results.

Different types of new data sources present different questions of an ethical nature – according to a range data characteristics, e.g. access to the data, content of personal information, quality issues in terms of suitability for the purpose of official statistics, the clarity of the methods to be applied in order to get statistical output etc.[15]

The *Table 1* included at the end of the paper represents a matrix, highlighting the most relevant principles to consider for each of the new data sources. Country examples for each of the types of new data sources used in official statistics are listed in Attachment 2.

Recommendations

1. NSOs should work with users and user communities to ensure the introduction of non-traditional and non-conventional data sources in official statistics continues to meet their needs and is well understood.
2. NSOs need to consider what additional engagement and information may be needed with users and user communities to support appropriate interpretation and use.
3. NSOs can use the matrix Table 1 to ensure that the relevant UNFPOS principles are considered sufficiently.

Further elaboration is now made using the types of new data sources used in official statistics and their linkage to respective UNFPOS which constitute further guidance on how to cover the mentioned issues in full compliance with the existing ethics.

B. Mapping against different types of non-traditional and non-conventional data sources

B.1   Mobile phone data

Mobile phone operators' systems generate a very large amount of data on the use of mobile communication, including location information. These data are mostly used for business and marketing purposes. However, the location data can be used for generating statistics about space-time movement of phones, needed for instance to supplement the tourism statistics.

The ethical issues at stake with this type of data concern the privacy of the data subjects (this non-conventional data source contains sensitive personal information), as well as the professional independence of national statistical offices that might be compromised while creating partnerships with mobile data providers in view of their equal treatment.

---

[15] Report on ethical guidelines, p. 6.

Principle 1 can be applied here, providing for ensuring the lack of biases connected with data collection in order to best meet the users' needs, to stay relevant and impartial. When it comes to further stages of the statistical production process, a major issue that can be identified during data processing is the lack of suitability of mobile phone data for statistical purposes. At the dissemination stage, because of differences in the definitions, the methodology concerning this type of data source can be complex. Here, Principle 3 is widely applicable, which requires a clear, transparent and understandable presentation of both official statistics and the corresponding metadata.

B.2   Data from smart electricity consumption meters

These data can be interesting for official statistics because they provide information on energy consumption, which can be beneficial for statistics on household consumption expenditure, consumer price indexes, environment statistics or statistics on energy consumption. The inclusion of this non-traditional data source into official statistics is well justified because of a considerable reduction in the burden on respondents, therefore the full application of Principle 5 can be observed here. A possible difficulty in securing access to a sufficient level of detail can be addressed using appropriate legal provisions, however the coverage of this data source should also be explored to assess if the statistics produced are representative and relevant. The smart meters' data are not intended for statistical purposes therefore the risk of discontinuity of the data source may also be present

B.3   Satellite imagery data

NSOs are exploring the possibility of using data from satellite imagery in official statistics[16]. They are expected to decrease burden on respondents, to improve timeliness, as well as to reduce survey costs[17]. They can also contribute to providing more disaggregated data. Satellite data are used to complement agriculture statistics along with other land cover/land use applications[18]. The increasingly widespread use of this kind of data is linked with relatively few concerns of ethical nature.

Data are mostly publicly available and they do not carry privacy concerns (or to a very limited extent). The major issue can be quality of statistical output that is directly related to the quality of satellite images, as well as methodology which, following Principle 3, should be clearly explained to the users. According to Principle 1, NSOs need to perform thorough research on the methods used to compile statistics based on satellite images to guarantee the quality of their final products. At this point we can observe the relevance of Principle 10

> "*Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries*"

Principle 10 provides for the joint efforts of international statistical and research communities targeted at finding common methodological solutions, sharing infrastructure, saving resources and taking advantages of synergies.

B.4   Social media data

Social Media can be used in at least three ways: As a subject of Official Statistics, e.g. use of Social media; to disseminate Official Statistics, thus reaching out to all kinds of users; and lastly as a source to compile Official Statistics. In what follows social media as a source (social media data) will be dealt with.

---

[16] https://unstats.un.org/bigdata/taskteams/satellite/UNGWG_Satellite_Task_Team_Report_WhiteCover.pdf
[17] http://eohandbook.com/
[18] https://ggim.un.org/documents/Paper_Land_cover_datasets_for_SDGs.pdf

Disseminated via the internet, social media data still represent an area for further exploration for official statistics. Adopting the form of messages, images, video or searches, these data are voluntarily submitted by users on the web. In several countries research is conducted to use social media to measure the level of well-being of societies (studies on happiness explore sentiment analysis).

The ethical concerns associated with these types of non-conventional data sources are related to privacy issues (there is a recurrent question whether the users of social networks should be notified that the information they post, although being public, will be used by statistical agencies) or to lack of access to data which sometimes has to be purchased from private owners. In this case an ethical question arises as to whether statistical offices should pay for data sources that are going to be transformed into public official statistics.

Processing this type of data source depends critically on the methodology adeptly to ensure a sufficient level of quality in the resulting official statistics. Therefore, strengthening research divisions in statistical offices is important if these type of data sources are to be used effectively.

Social media are also a vulnerable source of data when it comes to bias and manipulation. While disseminating statistics based on social media, it is important to accompany them with proper metadata and to describe them in an understandable way.

Again, while using this type of non-traditional data source to produce statistics, Principles 1, 2, 3, 6 and 9 could be applied here, to ensure relevance, to prevent loss of trust, to make them subject to comprehensive, comparable methodologies and standards, and to ensure users that their privacy is protected respectively.

The use of social media data should also be studied from the perspective of the fake news proliferation. Following Holan (2016), fake news can be defined as "invented material that has been cleverly manipulated so as to come across as reliable, journalistic reporting that may easily be spread online to a large audience that is willing to believe the stories and spread the message"[19].

In the context of social media, there is a danger of collecting data from users and making it available for advertisers who use it mainly to target advertisements. Another example of manipulating the truth is the existence of fake accounts, so called "bots" who may affect the factual image, hindering the quality of social media data to be used for statistical purposes.

Against this background, the ethical reference provided by the UNFPOS seems to be one of the most effective, as they promote official statistics rather than the raw social media data, thereby valuing more highly the production of reliable, comparable and high-quality data – those which meet international standards[20]. Elaborating on the arguments expressed during the conference "Truth in numbers: the role of data in a world of fact, fiction and everything in between"[31], firstly national statistical offices should be provided with the resources and infrastructure they need to support their roles as both a standard setter and coordinator across the national statistical system, and secondly it should be ensured by governments that there is no political interference in their national statistical systems, resulting in greater trust by citizens and reflecting national statistical offices as being "guardians of the facts"[21].

---

[19] Holan, A.D. (2016). "2016: Lie of the Year: Fake News," *Politifact*,December 13, 2016. Verified April 7, 2017: http://www.politifact.com/truth-o-meter/article/2016/dec/13/2016-lie-year-fake-news/ in Vincent F. Hendricks, Mads Vestergaard, *Reality Lost. Markets of Attention, Misinformation and Manipulation*, p.63., https://link.springer.com/content/pdf/10.1007%2F978-3-030-00813-0.pdf

[20] V. Hendricks, J. Jütting, *Trust in numbers? Why we need strong official statistics in an attention seeking society*, http://www.paris21.org/news-center/news/trust-numbers-why-we-need-strong-official-statistics-attention-seeking-society [31] 4 April 2018, Bern, Switzerland, http://www.paris21.org/news-center/events/conference-truth-numbers-role-data-worldfact-fiction-and-everything-between

[21] V. Hendricks, J. Jütting, *Trust in numbers? Why we need strong official statistics in an attention seeking society* [33] Report on ethical guidelines, p. 10.

B.5 Web-scraped data

Web scraping refers to a technique used for extracting data from websites. Official statistics uses webscraping techniques for example to collect prices of different goods from the internet and to use this type of data source as a supplement to the Consumer Price Index (CPI). If online prices replaced prices collected in a traditional way, costs of statistics could be considerably minimized. Similarly, to other non-traditional data sources, web-scraped data may raise legal issues (terms of use of websites differ across countries), they are not designed for statistical purposes, therefore methodological issues are also at stake. Once these data are made available, the statistics partially based on web-scraped data should— as in previously described cases—be accompanied by proper metadata.

B.6 Road traffic sensors data and passengers tracking sensors data

Vehicle detection loops, installed in pavements, or lasers can detect vehicles passing or arriving at a certain point, e.g. approaching a traffic light or in motorway traffic. or approaching a bus terminal. Normally, the data are stored in a central data warehouse of the responsible authority, e.g. the national transport agency.[33] Their use in official statistics is rather common. They serve to draw a picture of the number of vehicles, the speed with which they move, along with surveys for estimating commuting time. Passenger information collected by the sensors can be used to determine bus routes as well as monitor ridership to determine if service to certain areas should be increased or decreased. While using this type of non-conventional data source, the major issues are linked to quality. There are usually no privacy-related concerns. Therefore, the Principles that could be applied here refer to common methodologies, standards and definitions, as well to the proper explanation of statistics for the users.

B.7 Scanner and cash register data, e.g. from supermarkets

While purchasing a product in a supermarket, a record is being created by the scanner. These transaction records collected from many sellers are of great interest to statisticians. Used for the purposes of the Consumer Price Index (CPI) or complementing other domains of official statistics, such as business statistics or household expenditure, this non-traditional data source may outweigh the traditional collection methods mainly because of its lower cost.

Ethical issues in this respect concern i.e. establishing partnerships with private companies and can be linked with Principle 1—providing for relevance of statistics and professional independence of statisticians. Another question related to scanner data concerns methodological challenges (proper classifications of products, choosing appropriate index methodologies)—therefore the observance of Principles 3 and 9 can be considered a solution in the aforementioned cases.

B.8 CCTV data (Security / Surveillance videos), e.g. for citizen security purposes

Several countries / cities have decided to install camera's for citizen security purposes. While these may provide more comprehensive coverage of certain types of crime in certain areas, one has to be on the alert for both privacy issues (camera's capturing everybody: criminals and non-criminals alike) and quality issues (proper distribution of the cameras, so as to have complete coverage is a must).

C. Conclusion

Specific features of non-traditional data sources result in challenges for NSOs to comply with such principles as professional independence, access to sources, mandate for data collection, adequacy of resources, impartiality, objectivity and clarity of the methods used.

However, the technological development of today's world makes it impossible not to include these sources into official statistics' production processes – the outcome of which reinforces their complementarity to traditional data sources. As much as the NSOs should strive to build partnerships with non-official data producers and owners to make those data part of official statistical outputs, the cornerstone values of official statistics, such as quality, standards, and professional independence should never be neglected, nor compromised.

The relevance of the UNFPOS make them useful for the new technological reality. The proliferation of new data sources is a good opportunity to test and reinterpret the UNFPOS, and thereby, to confirm their universal character.

The mapping exercise that has been performed by the FOC FPOS confirmed an extensive applicability of the Principles to all aspects proper for new data sources. Moreover, owing to the fact that they constitute a basis for all data and statistical quality-related codes, charters, and frameworks throughout the world, it can be argued that they can respond to current challenges, be it the data revolution or the 2030 Agenda, with NSOs playing the leading and coordination role in the monitoring process.

Table 1 Mapping of the Fundamental Principles of Official Statistics against non-traditional and non-conventional data sources									
New data sources	Principle 1	Principle 2	Principle 3	Principle 4	Principle 5	Principle 6	Principle 7	Principle 9	Principle 10
DATA from	Impartiality principle, right for information (...) official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.	To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.	To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.	The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.	Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents	Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.	The laws, regulation and measures under which the statistical systems operate are to be made public.	The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.	Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.
Mobile phones	✓		✓	✓	✓	✓	✓		✓
Smart electricity consumption meters	✓		✓		✓	✓	✓		✓
Satellite imagery			✓		✓				✓
Social media	✓	✓	✓		✓	✓	✓	✓	✓
Web scraping activities			✓		✓				✓
Road traffic sensors			✓		✓				✓
Scanner and cash registers			✓	✓	✓	✓	✓	✓	✓
CCTV (security/ surveillance videos)			✓	✓	✓	✓	✓	✓	✓

26

Attachment 1: Examples of quality documents based on the UN Fundamental Principles of Official Statistics[22].

- The Fundamental Principles of Official Statistics
  https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx
- UNFPOS Implementation Guidelines https://unstats.un.org/unsd/dnss/gp/impguide.aspx
- European Statistics Code of Practice
  https://ec.europa.eu/eurostat/web/productscatalogues/-/KS-02-18-142
- Quality Assurance Framework of the European Statistical System (currently being reviewed)
  https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646
- Recommendation of the OECD council on good statistical practice
  http://www.oecd.org/statistics/good-practice-toolkit/Brochure-Good-Stat-Practices.pdf
- IMF Standards for Data Dissemination
  http://www.ins.tn/sites/default/files/methode/pdf/sdds_en_0.pdf
- ISI Declaration on Professional Ethics https://isi-web.org/index.php/news-from-isi/296declarationprofessionalethics-2010uk
- United Kingdom Code of Practice for Statistics which sets out the detailed principles and practices the UK statistics commits to, under three pillars of Trustworthiness, Quality and Value https://www.statisticsauthority.gov.uk/code-of-practice/
- CCSA Principles Governing International Statistical Activities
  https://unstats.un.org/unsd/ccsa/principles_stat_activities/
- Statistics Canada Quality Assurance Framework
  https://www150.statcan.gc.ca/n1/pub/12586-x/12-586-x2017001-eng.htm
- CARICOM'S Statistics Code of Practice
  https://caricom.org/documents/13410caricom_statistics_code_of_practice.pdf
- New Zealand's publication providing guidance to the National Statistical System about the highest priority official statistics (Tier 1).
  https://www.stats.govt.nz/assets/Uploads/Principles-and-protocols-for-producers-of-tier-1stats/principles-and-protocols-for-producers-of-tier-1-stats.pdf
- UN National Quality Assurance Frameworks Manual for Official Statistics
  https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3d-NQAF-Manual-E.pdf

- UN Global Working Group on Big Data
  https://unstats.un.org/bigdata/

---

[22] An extensive inventory of nationally and internationally developed data quality references is to be found at
https://unstats.un.org/unsd/methodology/dataquality/quality-references/

Attachment 2: Examples of non-conventional and non-traditional types of data in official statistics

1. Use of mobile phone data in official statistics
    - New Zealand used cell phone data to understand short-term population movements after the 2010 and 2011 Canterbury earthquakes. This was a special case for access to the data due to the natural disaster.[23]
    - UK made an estimation of commuter flows[24]
    - France has been conducting experimental research work on particular subjects (the evolution of the resident population throughout the day, mobility of people within local areas…) with the view of learning about the possibilities of using the CDR data for statistical studies. The experimental research work enables INSEE to be more familiar with the structure of the data and eventually to propose an aggregated format that would be interesting for statistical purposes.[25]

2. Use of smart meters data in official statistics
    - Estonia uses smart electricity meters data to produce electricity consumption statistics. A study on the feasibility of using smart meter data for producing information on electricity consumption was done with the view of using an opportunity to obtain reporting burden reduction on businesses. The output of the study is final energy consumption by economic activity, by region and monthly, quarterly and annual aggregation for businesses and final energy consumption by household characteristics as they are contained in household registers (size of dwelling, number of rooms and persons, etc.) by region and monthly, quarterly and annual aggregation. Another objective of the study was is to identify vacant dwellings and to verify real places of residence of households and related persons[26]. In the UK, the case was examined experimentally, using area-level summaries. The interests are twofold—validation of address data by cross referencing against aggregated counts of meter numbers and exploring the ability to identify areas with lots of vacant properties to save on the cost of field visits.[27]

3. Use of satellite imagery data in official statistics
    - Most statistical agencies use spatial data, if not satellite data, to map census boundaries. This is a pretty common use case. Many countries are also using satellite data in combination with other sources to produce SEEA Land Accounts:

---

[23] http://archive.stats.govt.nz/tools_and_services/earthquake-info-portal/using-cellphone-data-report.aspx
[24] the results of which can be found at:
https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrative datacensusresearchoutputs/populationcharacteristics/researchoutputsusingmobilephonedatatoestimatecommutingflows and https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/ onsmethodologyworkingpaperseriesnumber13comparingthedensityofmobilephonecelltowerswithpopulationesti mates
[25] For further details see the *Position paper on access to privately held data which are of public interest. Opening up new data sources for a new generation of official statistics – in light of the growing European Digital Single Market and the revision of the Public Sector Information Directive*, November 2017, p. 14 https://ec.europa.eu/eurostat/web/ess/-/access-toprivately-held-data. More information is available at : https://www.conference-service.com/NTTS2017/documents/agenda/data/abstracts/abstract_88.html and https://www.conferenceservice.com/NTTS2017/documents/agenda/data/abstracts/abstract_97.html
[26] See ESS position paper p. 18.
[27] Work to date is available at:
https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/programmesandprojects/theonsbigdataproject/analysinglo welectricityconsumptionusingdeccdata_tcm77-418326.pdf
https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/programmesandprojects/theonsbigdataproject/onssmartmet ercensusfeasibilitystudyfinalreportseptember2014v14_tcm77-408965.pdf
https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/programmesandprojects/theonsbigdataproject/comparingco untsofelectricitymetersandaddressesbypostcodeinenglandandwales.pdf

- Australian Bureau of Statistics produced an experimental land account including estimates of the value of land.[28]
- Statistics Canada has been developing Land Accounts to track land cover change since the 1990s. The current version is also used as the basis for experimental ecosystem accounting: https://www150.statcan.gc.ca/n1/pub/16-201-x/2013000/appendix-appendice1-eng.htm.
- Several countries in the ESCAP region are working on land accounts, but have not yet published: Fiji, Indonesia, Nepal, and Vanuatu.
- Statistics Poland is actively involved in activities aimed at including innovative statistical methods and techniques, covering the use of Big Data, in statistical production. The works are carried out as part of international projects (including the European Space Agency project), national ones (e.g."Gospostrateg", implemented by the National Center for Research and Development), as well as by means of ongoing negotiations with data owners (among others Google, Play (mobile phone operator), the Main Directorate of National Roads and Motorways). Areas in which solutions for advanced data analysis are developed, for the purposes of a real-time monitoring of the economy, which can be used to study selected aspects related to the single market policy and competitiveness, include among others agricultural research based on satellite imagery conducted by the European Space Agency. The use of satellite imagery technologies will allow to improve the identification and monitoring of agricultural crops and to develop methods for assessing the impact of extreme events such as: flood, drought, frosts, hail, etc. on the state of crops during the growing season, which cause direct repercussions in the economy (e.g. for prices of agricultural products, imports and exports, etc.);

4. Use of social media data in official statistics
- Statistics Netherlands uses social media messages to produce an alternative consumer confidence indicator. The research question was: Can we replicate the consumer confidence index by only using social media data, while reducing production time? A model has been built based on fitting characteristics derived from Facebook and Twitter messages.
- Istat produces an experimental social mood on economy index. The index was started by only using Twitter as a source, but further social media might be taken into consideration in the future.
- INEGI Mexico is working on the following Twitter-centered projects:
  - Mood of twitterers in Mexico – automatically measure and report the mood of twitterers, collecting more than 300 million twits. Process of sentiment classification, text normalization, machine learn algorithm, positivity quotient - Domestic tourism.
  - Mental health.
  - Mobility in Mexico City.
  - New agglomerations.
  - Consumer confidence.
  - Insecurity.
- Department of Statistics Malaysia (DOSM) uses online social media i.e. Twitter & Facebook for the statistical communication and dissemination of official statistics. Public Maturity Assessment on Official Statistics (PMAOS) is a tool used by DOSM to analyse sentiment analysis based on public opinion and reaction on the media. Data is presented in 3 different sentiment values – positive, negative and neutral. The analysis provides valuable input for DOSM to enhance its product and services.

5. Use of web-scraped data in official statistics
- Statistics Poland is carrying out works on development of methodology and implementation of modern systems for measuring the change index in retail prices – regarded as a measure of

---

[28] http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/1301.0Main+Features2822012

inflation – using and linking new data sources, among others data from information systems of retail chains (scanned data), data obtained directly from their owners (private companies), data on retail prices from the Internet (including using web scraping techniques), administrative data. As one of the key macroeconomic indicators, the index of changes in retail prices is used in analyzes and on-going monitoring of economic phenomena and the condition of the economy;

- Statistics Poland is carrying works on development of a methodology and tools for collecting data on the labor market in a real time. An analysis of the possibilities of acquiring data from corporate portals on which job vacancies are posted has been made, using web scraping techniques. Job offers are a key indicator in the analysis of the employment market, used among others by the European Central Bank and the European Commission as a measure of the demand on the labor market and the business cycle.

- Statistics Poland obtains information from websites using web scraping and text mining techniques, for the purpose of building a database of enterprises. Currently, entrepreneurs show very high activity on the Internet, their actions include looking for employees, selling products or sharing information about their new investments, hence the above research approach allows for a thorough real-time analysis of the situation on the enterprise market, including information on the possible flows of people, goods, services and cash.

- Statistics Portugal has been using web scraping since 2015 to collect prices for the Consumer Price Index (CPI) with the view of quality gains, such as cost and time reductions, and improved coverage and frequency. A technical solution was developed and tested for data extraction, storage, processing and analysis for an existent commercial website, comparing prices collected in a traditional manner. A traditional price collected from the outlets of the targeted company has been substituted with the same data found by the web scraper prototype. The example of Statistics Portugal is also the one of a successful strategic partnership with a private company.[29]

- UK Office for National Statistics (ONS) has done some early work on using web scraped information about consumer prices to derive experimental statistics [30]

- New Zealand has issued a number of papers on using web-scraped data for CPI measurement in New Zealand.[31]

- The Australian Bureau of Statistics (ABS) have collected retail prices for the Consumer Price Index (CPI) using web scrapers since May 2016. Web scrapers are currently programmed and maintained by ABS staff using Microsoft excel scripts (Visual Basic for Applications), collecting approximately 500,000 prices per week across over 50 retailers. The ABS has adopted a phased approach to implementing web scraped prices for production purposes. The first phase was implemented in June quarter 2017 where the ABS implemented web scraped data into the CPI, using an average calculated price of a product over a given period – essentially replacing field collected data with a web scraped price (ABS, 2017). That is, in the same way we currently price a specific product (whether in the field, online or via some other method), we choose a specific product in the web scraped time series to use as the basis for a price. With respect to the next phase of implementation, the ABS continues development work on text mining techniques (to form broader product definitions, especially for clothing) and price index methods (both bilateral and multilateral index methods) that maximise and automate the use of web scraped data.

- Department of Statistics Malaysia (DOSM) adopts web scraping techniques and crawls online price data from selected retailer's websites. The data is still being analysed and will be used as input for the computation Price Indices. The initiative is known as Price Intelligence.

---

[29] See: ESS position paper, p. 19.
[30] See
https://www.ons.gov.uk/file?uri=/aboutus/whatwedo/programmesandprojects/theonsbigdataproject/researchindi cesusingwebscrapeddatamay2016final002.pdf
[31] http://archive.stats.govt.nz/tools_and_services/newsletters/price-index-news/jan-18.aspx

6. Use of road traffic data in official statistics
- Statistics Poland is carrying research in the field of transport using data from sensors, such as the Automatic Ship Identification System (AIS) and Electronic Toll Collection System (viaTOLL), using stream processing which allows for data processing in a parallel and independent manner and for scaling the solution depending on the needs. In addition, a mechanism has been developed to enable quick analyses of real-time data from sensors together with historical data and data deriving from statistical surveys. Conducting transport analyses, using the above-mentioned data, allows to obtain new information on the volume of traffic, transport work, the volume of pollution emitted by transport, necessary to shape and monitor transport policy, including the selection of the most effective transport investments.
- The Department of Public Transportation (DPT) in Bermuda has tracked the number of passengers on public buses via a sensor fitted on each bus in the past. Upon the bus arriving at the central bus terminal, data was automatically downloaded once the bus sensor detected the terminal sensor. More than three million passenger journeys are typically recorded each year[32], indicating the large volume and frequency of events that were tracked automatically. This is a prime example of how the use of big data can be more accurate and efficient than the use of survey data which are subject to sampling errors and non-sampling errors, in addition to being costly and time consuming.

7. Use of scanner data in official statistics
- Statistics Poland is carrying out works on development of methodology and implementation of modern systems for measuring the change index in retail prices - regarded as a measure of inflation - using and linking new data sources, among others scanner data from information systems of retail chain.
- Statistics New Zealand uses scanner data for consumer electronics in our CPI measurement and continues to explore more opportunities.[33]
- Statistics Netherlands introduced supermarket scanner data into the CPI in June 2002 and in January 2010 they expanded the use of scanner data for the compilation of the Dutch CPI. The six chains for which scanner data were utilized at that time had a market share of around 50% and accounted for slightly more than 5% of the CPI-weight. In January 2013, scanner data fully replaced traditional price collection for supermarkets in the Dutch CPI.
- INSEE France has an objective of substituting scanner data owned by retailers for data collected in shops for industrial food products, household cleaning products and non-durable health and beauty products sold in food stores. They will be used to calculate the CPI in 2020. Since 2010, experimentations with 4 voluntary retailers representing 30 % of the market have been conducted[34].
- The Australian Bureau of Statistics (ABS) has used scanner (transactions) data in the production of the CPI since March quarter 2014. The ABS adopted a phased approach for implementation of scanner data, with the initial implementation replacing point-in-time prices for a sample of products (previously collected by field collectors) with a unit value (from scanner data). This implementation increased the coverage of scanner data to approximately 25% of the weight of the Australian CPI across several commodity classes including food, tobacco, and household non-durable products. From the December quarter 2017, the ABS implemented a new methodology (known as multilateral index methods) to compile

---

[32] Bermuda Government Department of Statistics (2014): *Bermuda Digest of Statistics,* p 77.
[33] http://archive.stats.govt.nz/browse_for_stats/economic_indicators/CPI_inflation/cpi-price-change-scannerdata.aspx
[34] See : ESS Position paper, p. 9. More information is available at https://www.INSEE.fr/en/statistiques/2912652

components of the Australian CPI using scanner data. This second phase of implementation overcome issues with traditional bilateral index methods and utilise a census of products available in these big datasets; uses expenditure data to weight products; and reduces data collection costs (ABS, 2016; ABS, 2017). In addition to its primary use in the Australian CPI, scanner data is utilised in other areas for both production and confrontation, including Retail Trade, National Accounts and Health surveys.

8. Use of CCTV data / Security videos
- The Bermuda Police Service has approximately 150 CCTV cameras across Bermuda which are monitored to assist with crime detection, 19 of which have license plate recognition capabilities. Incidents captured on the cameras are investigated before determining whether a crime has been committed and recorded as a statistic. Therefore, the big data captured by CCTV cameras indirectly leads to more comprehensive recording of crime statistics.

References

- Resolution adopted by the Economic and Social Council on 24 July 2013on the recommendation of the Statistical Commission (E/2013/24)2013/21. Fundamental Principles of Official Statistics

- United Nations Fundamental Principles of Official Statistics Implementation guidelines (Final draft, subject to editing, January 2015)

- Code of Principles in statistical cooperation of the Policy Group of Statistical Cooperation, Doc PGSC/2007/06

- UN Economic and Social Council, Resolutions. Supplement No. 1, Economic and Social Council official records, 50th session, 11–13 January and 26 April–21 May 1971. New York: UN, 1971.

- UN Economic and Social Council, Report of the Special Session of the Statistical Commission, New York, 11–15 April 1994, E/1994/29

- Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics amended by: regulation (EU) 2015/759 of the European Parliament and of the Council of 29 April 215

- Eurostat, European Statistics Code of Practice for the National Statistical Authorities and Eurostat (EU statistical authority). Adopted by the European Statistical System Committee 16th November 2017

- What does "big data" mean for official statistics?, United Nations Economic Commission for Europe, Conference Of European Statisticians, 10 March 2013What does big data mean for official statistics?, UNECE, CES, 10 March 2013

- Services concerning ethical, communicational, skills issues and methodological cooperation related to the use of Big Data in European statistics, Task 1. Ethical Review, Deliverable 1.3. Report on ethical guidelines June 2017.

- Abdulkadri Abdullahi, Evans Alecia, Ash Tanisha, An assessment of big data for official statistics in the Caribbean: Challenges and opportunities, January 2016

- Arora, Anil, Why are partnerships so important in the current context?, keynote speech from High-level Seminar on Strategic Partnerships in Official Statistics, Geneva, 11 April 2018

- Hendricks, Vincent.F., Jütting, Johannes, Trust in numbers? Why we need strong official statistics in an attention seeking society, http://www.paris21.org/news-center/news/trustnumbers-why-we-need-strong-official-statistics-attention-seeking-society

- Hendricks, Vincent F., Vestergaard, Mads, Reality Lost. Markets of Attention, Misinformation and Manipulation, p.63., https://link.springer.com/content/pdf/10.1007%2F978-3-030-00813-0.pdf

- Michaud, Sylvie TA2.14 Non-official Data: Challenges and Opportunities for NSOs, Canadian Experience, UN World Data Forum, Dubai 2018.

- Perucci, Francesca Implementing the Fundamental Principles in a transforming statistical system, 15th IAOS Conference, Abu Dhabi, UAE, 6-8 December 2016

- Rozkrut, Dominik, Capacity development for international cooperation in a modern statistics office in the face of new challenges – business as usual or value added? , keynote speech from the Conference of European Statistics Stakeholders, Bamberg, 18 October 2018.

- Statistical Capacity Building for Sustainable Development in : Developing the fundamental pillars necessary for modern national statistical systems/Statistical Journal of the IAOS 33 (2017) 895–909 895 DOI 10.3233/SJI-160331 IOS Pres.

- Position paper on access to privately held data which are of public interest. Opening up new data sources for a new generation of official statistics – in light of the growing European Digital Single Market and the revision of the Public Sector Information Directive, November 2017 https://ec.europa.eu/eurostat/web/ess/-/access-to-privately-held-data

- Report by the UN Big Data Global Working Group Satellite Imagery and Geo-Spatial Data Task Team: https://unstats.un.org/bigdata/taskteams/satellite/

- Australian Bureau of Statistics 1352.0.55.144 - Research Paper: Methodological Approaches for Utilising Satellite Imagery to Estimate Official Crop Area Statistics (Methodology Advisory Committee) http://abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1352.0.55.144Sep%202014?OpenDocument

- Australian Bureau of Statistics (ABS). Making greater use of transactions data to compile the consumer price index. Information Paper 6401.0.60.003, Canberra: ABS; 2016.

- Australian Bureau of Statistics (ABS). Feature Article: The Australian CPI: A Contemporary Measure of Household Inflation. Cat. no. 6401.0, Canberra: ABS; 2017.

- Australian Bureau of Statistics (ABS). An implementation plan to maximise the use of transactions data in the CPI", Information paper 6401.0.60.004, Canberra: ABS; 2017.