

LEGAL ASPECTS RELATED TO ENHANCE THE COOPERATION WITH DATA HOLDERS: THE CASE OF SPAIN

Gregorio Izquierdo Llanes, President of INE-Spain, gregorio.izquierdo@ine.es.

Antonio M. Salcedo, President Office, Director antonio.salcedo.galiano@ine.es.

Yolanda Gómez, President Office, Legal Advisor, yolanda.gomez.menchon@ine.es.

Abstract

The concept of cooperation is related to coordination in Regulation 223/2009 and in the Code of Practice of European Statistics. Cooperation is understood as horizontal tool between national statistical authorities and other public entities or between statistical authorities and the scientific community, or even among statistical authorities of different Countries (International cooperation). However, the ESS Quality Declaration adopted in 2016 introduce a new meaning stating that “We consistently strive to minimise the burden on our respondents, cultivate good cooperation with data providers and cooperate closely with stakeholder groups including the scientific communities”. The first question that comes to mind is if the ESS is moving from the “traditional” meaning of cooperation among public data providers (the owners of the administrative data) to an enlarged cooperation with the private sector. In particular, focusing in a very hot issue like Big Data, we analyse in this article some issues such as how this cooperation could be, if there is a need for legal support for cooperation with data holders or, if so, which type of legal acts could be more useful.

Keywords: European Statistics, ESS Quality Declaration, Big Data, Cooperation, institutional environment, legal framework.

1. Introduction

The European Statistical System (ESS) has been sending the message that statistics and the way we produce them are changing since august 2009 when the European Commission (Eurostat) published its Communication on the production method of EU statistics: a vision for the next decade (the Vision). In the subsequent nine years the ESS has been working in different areas in the direction pointed out in the Vision and in an updated version (Vision 2020). One of these areas is related to the use of Big Data to produce official statistics.

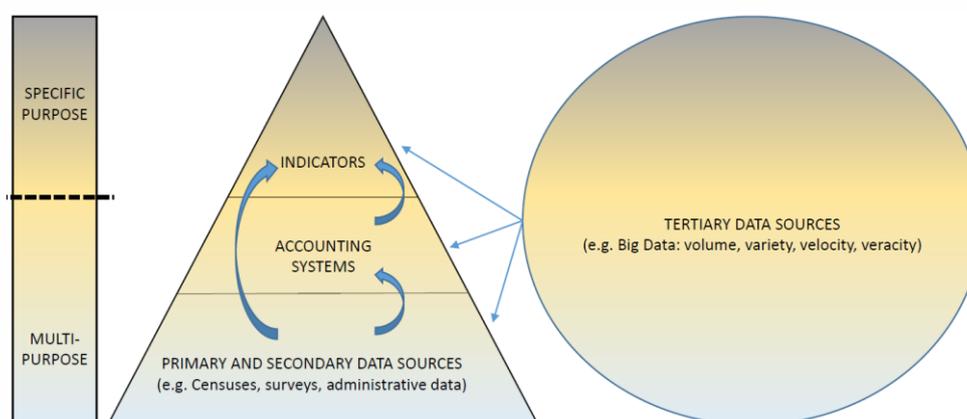
This innovative approach has many collateral issues that must be addressed before testing the real added value of the use of these data. In this article we introduce the

current situation related to the use of Big Data for producing official statistics in the ESS in general in order to make an analysis of the main legal and practical issues related to this new data source, taking into account the Spanish experience.

2. Big data and official statistics

The European statistical programme specifies the programming framework for the development, production and dissemination of European statistics. This programme provides the classical approach to the infrastructure of statistical information, based on the pyramid of the statistical information:

Fig 1: The classical pyramid of statistical information.



Source: Prepared by the authors based on the European Statistical Programme 2013-2017-2020

The base of the pyramid is formed by primary and secondary data sources, that is, data collected directly by the NSIs usually by sampling surveys or census, or data obtained from administrative sources. These primary and secondary data sources are, beside the accounting systems and the set of indicators, the basis of the core statistical production that are used at the decision-making process.

The use of Big Data in official statistics implies the introduction of a “third” data source that, a priori, is not originally designed for producing statistics because of many different reasons, such as concepts and classifications not internationally harmonized, unstructured data, an usual lack of metadata, etc.. Therefore, it requires an important effort to pre-process the information. The taxonomy developed by the United Nations Conference of European Statisticians (UNECE) gives us a first contact to the different types of Big Data.

Table 1: Types of Big Data and examples of statistical applications in Member States.

UNECE Classification of Types of Big Data	Examples of statistical activities in the Commission and Member States (*)
---	--

<p>1. Social networks 1100. Social Networks: Facebook, Twitter, Tumblr etc. 1200. Blogs and comments 1300. Personal documents 1400. Pictures: Instagram, Flickr, Picasa etc. 1500. Videos: YouTube etc. 1600. Internet searches 1700. Mobile data content: text messages 1800. User-generated maps 1900. E-Mail</p>	<p>- Exploring the use of social media messages for economic indicators (NL). - UNECE Wiki Project: studying correlations between wikistats and official statistics (IE). - Using Google Trends for "now-casting" unemployment (ESTAT)</p>
<p>2. Traditional business systems 21. Data produced by Public Agencies 22. Data produced by businesses</p>	<p>- Scanner data for CPI/HICP (IT). - Credit and debit cards' data in National Accounts (PT).</p>
<p>3. Internet of things 31. Data from sensors (fixed or mobile) 32. Data from computer systems</p>	<p>- Use of smart electricity meters to produce electricity consumption statistics (EE). - Big Data for the estimation of origin-destination mobility matrices (ES).</p>

(*) Source: Prepared by the authors based on the information from the UN Big Data Project Inventory

The starting point to analyse the issue of the use of Big Data in the ESS is the Scheveningen Memorandum elaborated at the DGINS Conference and adopted by the 18th ESSC in September 2013. In the Memorandum, the ESS acknowledge that Big Data represent new opportunities and challenges for Official Statistics, and therefore encourage the European Statistical System and its partners to effectively examine the potential of Big Data sources and agree on the importance of following up the implementation of the memorandum by adopting an ESS action plan and roadmap by mid-2014 that should be further integrated into the Statistical Annual Work Programmes of Eurostat.

The memorandum recognises that the implications of Big Data for legislation especially with regard to data protection and personal rights (e.g. access to Big Data sources held by third parties) should be properly addressed as a matter of priority in a coordinated manner. It also acknowledge that the multidisciplinary character of Big Data requires synergies and partnerships to be effectively built with experts and stakeholders from various domains including government, academics and owners of private data sources.

The Road Map on Big Data adopted in 2014 by the ESSC stressed that legal and ethical issues related to big data access and use, as well as the development of the appropriate skills, are of strategic importance and should be given high priority. Since then, the ESS has been working on it as one of the Vision Project (BIGD) with the purpose of identifying, analysing and solving issues in horizontal areas or topics as well as investigating and developing future business models for statistical data production related to specific data sources or statistical products. As part of this project

the ESSnet Big Data has started in February 2016 and is to run for 28 months until May 2018.

The work of this ESSnet through different studies and pilot is illustrative of how Big Data could be used for statistical purposes. Far from following the structure set up by the European Statistical Programme 2013 to 2017 enlarged until 2020, the work started with the different sources and tools to look for practical uses in statistics. The ESSnet project consists of 10 work packages, eight of these are content-oriented, while the other two, Coordination and Dissemination, support the overall project. We can see, for example, that the experimental work has already been done for producing early estimates of certain macroeconomic indicators (GDP and short term turnover indices), for producing estimates relevant to population, tourism, and agriculture by combining big data, administrative sources and traditional surveys.

3. Legal and practical issues.

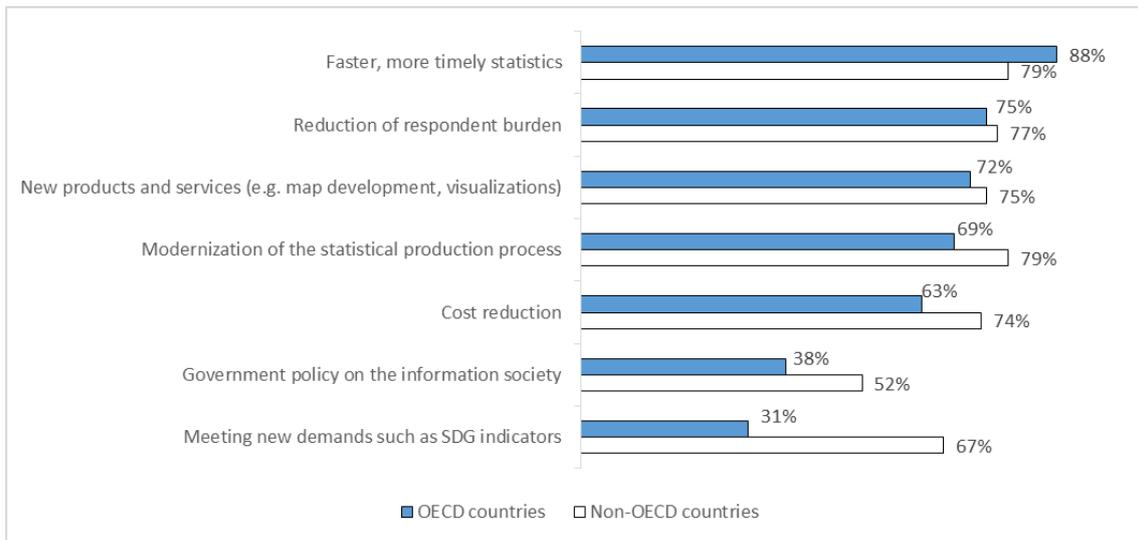
The production of official statistics responds to the highest quality standards according to the European Statistics Code of Practice. To this aim, the use of Big Data arises as a good opportunity to improve the classical approach to produce official statistics, reducing the response burden and increasing the timeliness of the statistics while maintaining an apparently moderate increase of the associated costs in the medium/long term. In addition, the inclusion of Big Data for official statistics would enable the production of new statistics and the modernization of the NSIs through the most innovative IT tools to process this kind of data.

In short, the benefits of the use of big data can be explained from two perspectives: From one hand, outward-looking, we could expect more adequate and flexible response to user needs, a wider range of statistical products and services (without increasing burden), a better understanding of quality aspects of new sources and a higher temporal and spatial resolution. From the other hand, inward-looking, we will acquire new competences for NSIs, increase efficiency in producing statistics while remaining key players for statistical information (Wirthmann, 2016).

The use of Big Data is associated with a paradigm shift for which many statistical offices are not yet fully prepared. The study carried out in 2016 by the United Nations Statistical Commission reveals that the majority of the OECD statistical offices considers as main reasons for the use of Big Data, the production of faster and more

timely statistics, the reduction of the response burdens and the production of new products and services:

Fig 2: Main reasons for and business benefits of using big data.



Source: Prepared by the authors based on the Report of the Global Working Group on Big Data for Official Statistics (Forty-seventh session, Statistical Commission)

The projects for including the use of Big Data for official statistics have been focused in different aspects of the generic statistics business production model, trying to solve all the arising collateral problems. As an example, in the collection phase, one of the main issue is how we will get the data: Should we have access to the raw data by any valid physical support? Should we have access online or may be in the data holder premises or through third parties? Or should we have to pay for the data or the processing of the data?

Table 2: Official statistics traditional approach and big data paradigm shift.

Phases of the Generic Statistics Business Process Model (GSBPM)	Big Data processes (*)
<ul style="list-style-type: none"> • Specify Needs • Design • Build • Collect • Process • Analyse • Disseminate • Evaluate 	<ul style="list-style-type: none"> • Data Management <ul style="list-style-type: none"> - Acquisition and recording. - Extraction, cleaning and annotation - Integration, aggregation and representation. • Analytics <ul style="list-style-type: none"> - Modelling and analysis. - Interpretation.

(*) Prepared by the authors based on the classification made by Gandomi & Haider (2015)

We can find some other issues at the dissemination phase, e.g., due to the fact that one of the reasons why the data holders are reticent to give access to the data is because it can be part of their business, of their asset, thus, they do not want anybody

to publish any results obtained with these data. We have mentioned these examples but in real life the use of Big Data may have a real impact in all the different phases of the GSBPM¹ model.

Therefore, the official statistical existing model contrasts with the paradigm shift that is required by the processing of Big Data requires. Even if the data is already there it is necessary to find the way of properly collect and process them exploring new tools and know-how and, finally, validate the results.

As we have introduced, to succeed in the analyses of the methodologies and uses of Big Data the first pre-condition is to have access to the data². This situation lead us to analyse in deep the Spanish statistical legislation. Article 10 of the national statistical law³ states clearly that *“1. The statistical services may request data from all natural and legal persons, national and foreign, resident in Spain. 2. All natural and legal persons that provide data, whether their collaboration is mandatory or voluntary, must answer truthfully, accurately, completely and within the deadline to the questions ordered in due form by the statistical services”*. Moreover, all European statistics are automatically included in the National Statistical Plan and all statistics included in the National Statistical Plan are mandatory. Article 4 establishes that the collection of data for statistical purposes shall conform to the principles of secrecy, transparency, specialty and proportionality.

It's worth mentioning that this is not only the case in Spain, since there are other cases at EU level showing that the national and EU statistical legislation confers enough legal power to the statistical offices to request the access to those data, but a different question is the associated complexity of the access (Salgado, 2017), we are forced to open cooperation ways as we did in the past and still we do with the administrative data holders.

Regarding the ethical issues and the right of collecting the data, first of all it's essential to clarify who is the owner of the data and the associated rights. According to our national data protection law and in line with the General Data Protection Regulation (GDPR), the data holder is the responsible of the data, the controller, and the one who has to protect it, as a consequence, the issue about who is the ownership (the data

¹ Generic Statistical Business Process Model.

² This was the aim of the WP5 Mobile phone data of the ESSnet in which Spain has been participating.

³ Ley 12/1989 de la Función Estadística Pública

subject or the controller) is not so relevant for processing the data of natural persons for official statistics, while is still important when the data subject is a legal person. Different is the legal collateral issue if those data are subject to some copyright and/or commercial property rights (that in the EU is part of the intellectual property), because here we should provide justification that there is not a real competition or risk on those rights as we are a non-profit organisation acting in a public interest, but it will depends on the different national law. In addition, some specific data has special legislation that might conflicts with the statistical law. As we see, another issue that we have to take into account when diving into ethical issues is that they are closely linked with legal issues (i.e. rights, independence, mandate for collection, confidentiality, transparency...).

Nonetheless, the ethical issues have to be analysed from the institutional or professional perspective. In official statistics this is related to the principles or standards that the statistical authorities shall follow when producing official statistics. Therefore, we have to focus on the principles developed in the European Statistics Code of Practices. Some authors have analysed the situation in detail in connection to the principles of Code making some improvement proposals (Rutkauskiene, 2017). As regards the statistical confidentiality, for example, we believe we can that there is enough margin of manoeuvre since recital 50 of the GDPR states that “The processing of personal data for purposes other than those for which the personal data were initially collected should be allowed only where the processing is compatible with the purposes for which the personal data were initially collected. (...) Further processing for (...) statistical purposes should be considered to be compatible lawful processing operations”, and this is ruled also in article 5.1.b. As regards the duty of information the GDPR introduces also an exception for statistical purposes “it is not necessary to impose the obligation to provide information where (...) the provision of information to the data subject proves to be impossible or would involve a disproportionate effort. The latter could in particular be the case where processing is carried out for (...) statistical purposes”. Moreover, we consider that the statistical confidentiality is applicable independently of the kind of data⁴ because Regulation 223/2009 define “statistical confidentiality” as the protection of confidential data related to single statistical units

⁴ Rutkauskiene considers that when the observation unit is a car or a mobile phone or a sum of purchases etc., it cannot be associated with individuals (natural or legal person). With this argument, as they are not individual data, the statistical principle of confidentiality is not applicable. (Rutkauskiene, 2017)

which are obtained directly for statistical purposes *or indirectly from administrative or other sources* and statistical unit is defined as the basic observation unit, namely a natural person, a household, an economic operator and other undertakings, referred to by the data.

From the ethical and legal perspective, if the processing of personal data is lawful for official statistical purposes without the need, or even the particular knowledge, of the data subject and they will be protected by the statistical confidentiality rules, then, the private data holders should not deny the access to data of natural persons for privacy reasons. As regards data of legal persons for which specific privacy rules may apply, leaving the commercial reasons apart, those privacy bilateral rules should not be put in front of the statistical law to deny access to Big Data.

Some of these issues have been addressed by the ESS updating the Code of Practices in 2017 and the Quality Assurance Framework is being also adapted to this amendments. In particular, the code has added, as part of principle 8, a new need of cooperation with private data holders together with the cooperation with researches and universities. In fact, The Code itself and its further amendments have been boosted by the need to enhance trust in official statistics. For the use of Big Data the message here is clear, the ESS wants everybody to know that we are open to cooperate, to respect security and privacy of data while ensuring a sound methodology and high quality. But how should we cooperate?

4. Cooperation with data holders

4.1. From collaboration to cooperation in the ESS

In general collaboration and cooperation are almost synonyms but in practice, at least in official statistics, the evolution of the ESS shows that cooperation implies a greater compromise for the parties involved.

In the nineties, the first European statistical law (Regulation 322/1997 on Community Statistics) recognizes the need to reinforce collaboration and coordination between statistical authorities to ensure the feasibility, coherence and comparability of Community statistics. This “collaboration and coordination” become in 2009, under Regulation 223/2009 on European Statistics, *in coordination, collaboration and cooperation*. Regulation 223/2009 recognised the need of reinforcing cooperation and

coordination between the authorities in a more systematic and organised manner. The members of the European Statistical System should create synergies through collaborative networks. As regards the need of cooperation the statistical law moved from the mention of the needed cooperation with the European Monetary Institute (the predecessor of the European Central Bank) to a wider cooperation. The statistical law includes the external cooperation, but also the need of internal cooperation between Member States and the Commission to reach common ESS position at international fora or to reach a coordinated release and revision policy when using the European approach to statistics.

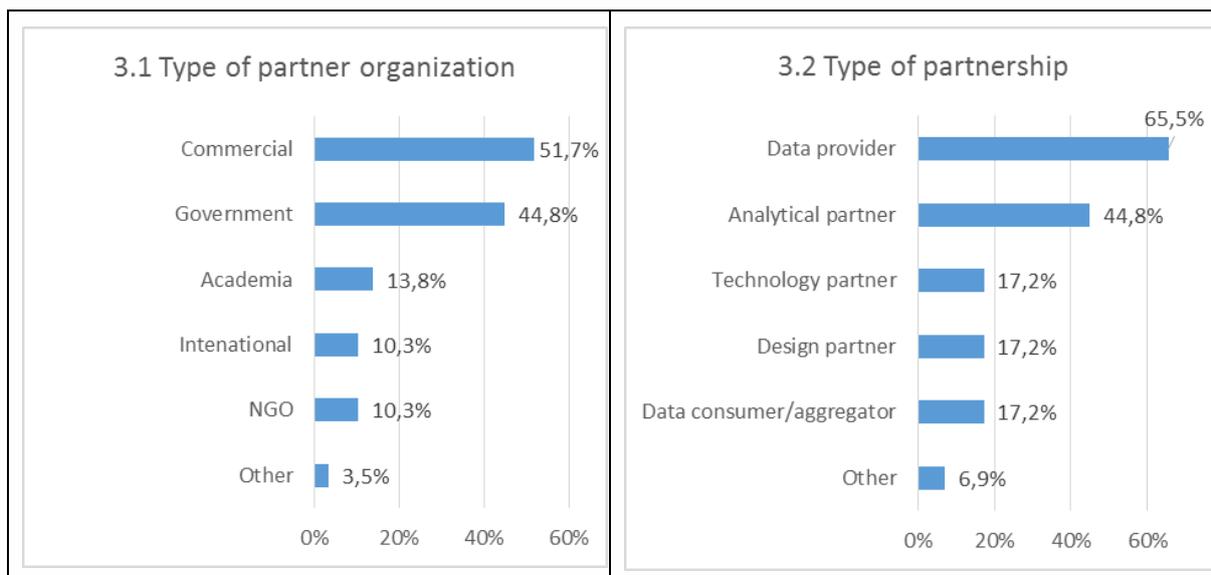
The European Statistics Code of Practice enhance this internal and external cooperation. However, the ESS Quality Declaration adopted in 2016 introduced a new meaning stating that *“We consistently strive to minimise the burden on our respondents, cultivate good cooperation with data providers and cooperate closely with stakeholder groups including the scientific communities”*. In line with this Declaration and with the result of the pilot studies carried out under the Vision Project BIGD the ESSC adopted the revised Code of Practices in order to stress the cooperation with the data holders, as mentioned above.

4.2. A real partnership?

The most recent studies talk of “Partnerships”. It is interesting the classification made by the UN Global Working Group on Big Data for Official Statistics Task Team on Access and Partnerships in 2015. First, the Group states that “a partnership can be defined in legal or non-legal terms, i.e. “a legal relation existing between two or more persons contractually associated as joint principals in a business “, or “a relationship resembling a legal partnership and usually involving close cooperation between parties having specified and joint rights and responsibilities”.

In the wider context of Big Data for official statistics, partnerships include the relationship of two or more organizations conducting actual collaborations without necessarily resembling a legal partnership”.

Fig. 3: Partnerships with other organizations or data providers.



Source: Prepared by the authors based on the results of the UNSD/UNECE Survey on organizational context and individual projects of Big Data (2015)

This issue is related with the available instruments to enhance cooperation with data holders. The literature is unanimous recommending that the cooperation tool, whatever the form it takes, shall address all the possible problems and concerns, shall balance the benefits and risks of both parties (ownership, confidentiality, modalities of access, quality, methodology, dissemination, etc.). Presented like this, it has the appearance of a synallagmatic relationship ruled by a private contract or a gentlemen agreement and the questions here are: is this is a real cooperation, a real partnership? Are we working to the same aim? Should the public administration be on equal footing as private companies?

The joint steps done by the ESS shows that the NSIs cannot be on equal footing as private companies to produce statistics. In November 2017, the ESS has issued the “Position paper on access to privately held data which are of public interest” in which the ESS “invites the European Commission (...) to include in a future proposal amending the PSI Directive, relevant provisions that would establish a general principle of access to privately-held data which are of public interest, while addressing in broad terms key aspects that would ensure that such access is effective and sustainable”. Even if the result of the public consultation was that “88% of the 205 respondents supported the access to data from private sector entities for reasons of public interest and that specific legal measures should be put in place. The Commission has decided to start with soft law publishing a “Guidance on sharing private sector data in the European Data Economy” as part as the Commission Communication “Towards a

Common European Data Space”, leaving its regulation in the PSI Directive for the medium term.

The first conclusion is that the ESS agreed that we need a common legal basis in the EU. This is the first shot that push Member States to implement the Directive specifying in more detail the access. Thus, if we succeed in having this legal basis in the medium term, do we still need cooperation with private data holders or, better said, data tenants? The answer has to be positive, but it is important that we let them know, through the PSI Directive or the implementing national law, that we do not pursue private interest but public one. We have the duty to produce high quality official statistics subjected to strict ethical principles and sound methodologies. From that starting point we could build a cooperation framework in which the reciprocity shall be based on a mutual benefit but not from a commercial perspective, the NSIs should not pay for acceding the data. There are different incentives that we could offer to the private data holders. Klein and Verhulst identify different incentives such as given them access to other data sources (while complying with the statistical confidentiality, it is understood), tap data analytical skills available outside their own company, enhance a firm’s corporate image and reputation, potentially attracting new users and customers, just by corporate social responsibility or philanthropy...(Klein, T and Verhulst, S, 2017). One important element is to offer security to the data tenant about the data shared and explaining clearly in the contract/agreement the dissemination policy. In the Spanish experience we have found that some companies doesn’t mind to share part of the data to research purposes but not to produce final statistics, they do not want to allow the dissemination of the results which lead us to a death end. This is linked to the issue of the risk of volatility that has the use of Big Data for producing official statistics. We must assure a long-term cooperation although we could never prevent the end of the business of the data tenant and, consequently, the termination of the collection of the Big Data source. For this reason, as the experts in this field recommends, we need to use Big Data in a complementary way along with other traditional sources.

The Commission Communication and the “Guidance on sharing private sector data in the European Data Economy” recognises the principles that must guide these new type of partnerships: Proportionality in the use of private sector data, purpose limitation, do-no-harm, conditions for data re-use, mitigate limitations of private sector data and

transparency and social participation and includes also some possible models of cooperation (data donorship, prizes, partnerships, intermediaries or civic data sharing).

4. Conclusions

In this paper we have reviewed some important issues related with the use of Big Data. Our aim has been to make a brief analyse of some ethical and legal obstacles found to access Big Data from private data tenants. In line with the latest developments in the ESS we claim for the need of a minimum legal basis that help cooperation with private companies. Nevertheless, this is essential but is not enough.

In the post-truth era it is critical for statisticians to understand more deeply how information is processed by individuals and groups. This affects the ability to communicate correctly, but also the very design of the statistical products and services provided. We are facing the risk that communication impacts of official statistics could be hampered by lack of trust and credibility on the part of the users (Badacci and Pelagalli, 2017).

There is a clear need to strength the cooperation with data tenants on the basis of solid legal rules, this will be essential to improve the classical approach to produce official statistics by increasing the quality, in particular timeliness and accuracy, and reducing the response burden.

The ESS is moving in the right direction but we also need additional support to increase the cooperation among all EU institutions and among the Member States as the basis to produce high-quality statistics using new production methods to support decision-making based on evidence and meet the current and future requirements of our users.

5. References

Agrawal, D et al (2012). Challenges and Opportunities with Big Data: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. Available at: <https://cra.org/ccc/resources/ccc-led-whitepapers/>

Baldacci, Emanuele and Pelagalli, Felicia (2017): Communication of statistics in post-truth society: the good, the bad and the ugly. Eurostat?

Cornelia, L et al. (2017): Big Data: Potential, challenges and statistical implications. IMF Staff Discussion Note (SDN/17/06).

European Commission (2018): Communication from the Commission to the EP, the Council, the EESC and the CR "Towards a common European data space" and the "Guidance on sharing private sector data in the European data economy" (SWD(2018) 125 final)

European Parliament (2018): Fact Sheets on the European Union on Intellectual, industrial and commercial property.

European Statistical System (2013): Scheveningen Memorandum.

European Statistical System (2017): European Statistical Code of Practices. http://ec.europa.eu/eurostat/documents/64157/4392716/Revised_CoP_Nov_2017.pdf

European Statistical System November (2017): Position paper on access to privately held data which are of public interest.

Eurostat, 2005: Communication from the Commission to the European Parliament and to the Council on the independence, integrity and accountability of the national and Community statistical authorities (COM(2005) 217 final)

European Statistical System November (2017): Position paper on for Access to Data from Private Organizations for Official Statistics

Gandomi, Amir and Haider, Murtaza (2015): Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35 (2015) 137–144.

ISI (2010): ISI Declaration on Professional Ethics

Klein, Thilo and Verhulst, Stefaan (2017): Access to New Data Sources for Statistics: Business Models and Incentives for the Corporate Sector. PARIS21 works.

Landefeld, Steve (2014): Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues. Discussion Paper on the International Conference on Big Data for Official Statistics, Beijing, China, 28 – 30 Oct 2014.

Ley 12/1989, de 9 de mayo, de la Función Estadística Pública.

Radermacher, Walter (2017): Conference "Power from statistics: delivering the evidence of tomorrow". Available at: <http://powerfromstatistics.eu/OR/PfS-OutlookReport-Radermacher.pdf>

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Regulation 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics

Regulation 322/1999 of the Council on Community Statistics.

Rutkauskiene, Alma (2017): Report on the review of ethics related to big data. SOGETI, Eurostat.

Salgado, David (2017): Big Data en la Estadística Pública: Retos ante los primeros pasos. Economía industrial, N° 405, 2017 (Ejemplar dedicado a: Nuevas tecnologías digitales), págs. 121-129

Salgado, David et al (2017). Guidelines for the access to mobile phone data within the ESS. ESSnet on Big Data WP5 Deliverable 1.2. https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/6/65/WP5.Deliverable_1.2.pdf

Struijs, Peter (2016): Big data for official statistics. Eustat, Seminario internacional de estadística. Available at: http://www.eustat.eus/productosServicios/datos/58_Big_Data_for_Official_Statistics_Peter_Struijs.pdf

United Nations (2014): Fundamental Principles of Official Statistics.

United Nations (2015): Draft Good Practices for data access and partnerships. (TTAP)

United Nations (2016): Recommendations for access to data from private organizations for Official Statistics. Working Group on Big Data. Dublin.

United Nations, Economic Commission for Europe, Conference of European Statisticians, 62 plenary session (2014): In-depth review of big data.

Wirthmann, Albrecht (2015): Structuring risks and solutions in the use of big data sources for producing official statistics – Analysis based on a quality framework. United Nations, ECE (Workshop on Statistical Data Collection: Riding the data deluge).

Wirthmann, Albrecht (2016): The Data Revolution and Official Statistics. Eurostat, Workshop on Partnerships. Luxembourg.