

Eurostat staff publications with abstracts – author-year style, year - name - title order

Book chapters

Bujnowska, A. (2019). “Access to European Statistical System microdata”. In: *Data-Driven Policy Impact Evaluation – How Access to Microdata is Transforming Policy Design*. Ed. by N. Crato and P. Paruolo. Springer, pp. 87–99. DOI: [10.1007/978-3-319-78461-8](https://doi.org/10.1007/978-3-319-78461-8). URL: <https://www.springer.com/gp/book/9783319784601>.

Bach, F. (2018). “Statistical disclosure control in geospatial data: The 2021 EU Census example”. In: *Service-Oriented Mapping – Changing Paradigm in Map Production and Geoinformation Management*. Ed. by J. Döllner, M. Jobst, and P. Schmitz. Lecture Notes in Geoinformation and Cartography. Springer. Chap. 18, pp. 365–384. DOI: [10.1007/978-3-319-72434-8_18](https://doi.org/10.1007/978-3-319-72434-8_18). URL: <https://link.springer.com/content/pdf/10.1007/978-3-319-72434-8.pdf>.

This chapter outlines challenges and modern approaches in statistical disclosure control of official high-resolution population data on the example of the EU census rounds 2011 and 2021, where a particular focus is on the European 1 km grid outputs derived from these censuses. After a general introduction to the topic and experiences from 2011, the recommended protection methods for geospatial data in the planned 2021 census 1 km grids are discussed in detail.

Buono, D., E. Infante, and G. L. Mazzi (2018). “Short versus long time series: An empirical analysis”. In: *Handbook on Seasonal Adjustment*. Publications Office of the European Union. Chap. 25, pp. 669–680. DOI: [10.2785/941452](https://doi.org/10.2785/941452). URL: <https://ec.europa.eu/eurostat/documents/3859598/8939616/KS-GQ-18-001-EN-N.pdf>.

Most of the literature on business cycle analysis relies, as input, on the seasonally adjusted (SA) data of the main economic indicators. The rationale is that the seasonal frequencies are different from the frequencies of the cycles, then seasonal movements do not carry useful information, moreover they can hide the information on the frequencies of interest in business cycle analysis. This idea is coherent with the literature on SA that rests on the hypothesis of orthogonality among the seasonal and the others components. For several reasons this hypothesis can fail and an interaction between seasonal and business cycles can arise. This work address the plausibility of this hypothesis and a first study on the effect that different seasonal adjustment algorithms can have on the business cycle analysis due to their different ability in separating the seasonal from the other frequencies. Empirically the evidence or the presence of interactions can be hardly detectable. There are several reasons: components are unobservable as well as their connections, consequently. Moreover, the series are often characterized by instability, and/or evolutionary behaviour of the components. Aim of this work can be summarized in: empirical investigation of the effects of a variety of SA methods on two aspects: the cyclical shape of the series and the turning point dating. The focus is on the

growth cycle. The approach will be historical, and then no real time exercises is run.

- Gatto, R., D. Ladiray, and G. L. Mazzi (2018). “The effect of alternative seasonal adjustment methods on business cycle analysis”. In: *Handbook on Seasonal Adjustment*. Publications Office of the European Union. Chap. 23, pp. 629–654. DOI: [10 . 2785 / 941452](https://doi.org/10.2785/941452). URL: <https://ec.europa.eu/eurostat/documents/3859598/8939616/KS-GQ-18-001-EN-N.pdf>.

Seasonal adjustment procedures are usually designed for being applied on sufficiently long time series in order to obtain good quality results. This is due both to technical reasons such as the properties of the symmetric filters used and to non-technical ones such as the fact that, over a sufficiently long time period, components can be better identified and separated; consequently, the seasonal component can be more precisely estimated and eventually removed. In addition, long time series are required in order to read properly the statistics of the seasonality tests. Furthermore, on short time series the seasonality tests might be less robust. In official statistics, available time series associated to statistical indicators are often relatively short or subject to some shortening processes. This seems to contradict one of the main quality dimensions of statistics which is the coverage, but, at the same time, official statistics need to be continuously improved to better reflect the socio-economic structure. This process unavoidably leads to regularly adapting existing official statistics to evolving socio-economic structure. Furthermore better reflecting the current socio-economic situation can also require the development and the statistical compilation of new indicators, which at least in a first phase will cover only a limited time span. These two processes imply, at least temporarily, the availability of short time series associated to the statistical indicators.

- Eiselt, B. (2016). “LUCAS-Erhebung: Bodenbedeckung und Bodennutzung in der EU”. In: *Flächennutzungsmonitoring VIII. Flächensparen – Ökosystemleistungen – Handlungsstrategien*. Ed. by G. Meinel et al. URL: <http://slub.qucosa.de/api/qucosa/%3A16825/attachment/ATT-0/>.

- Chiappero-Martinetti, E. and A. Sabadash (2014). “Integrating human capital and human capabilities in understanding the value of education”. In: *The Capability Approach: From Theory to Practice*. Ed. by S. Ibrahim and M. Tiwari. MPRA Paper 61800. London, UK: Palgrave Macmillan. Chap. 9, pp. 206–230. DOI: [10 . 1057 / 9781137001436 _ 9](https://doi.org/10.1057/9781137001436_9). URL: <https://link.springer.com/content/pdf/10.1057%2F9781137001436.pdf>.

The aim of this chapter is to investigate the possibility of combining human capital theory (HCT) and the capability approach (CA) in order to better understand and measure both the instrumental and the intrinsic values of education for individuals, and to trace its relative spillover effects on societies. HCT, pioneered by Schultz and Becker in the early 1970s, has since become an important part of the debate on economic growth and development. Recently, HCT has been criticised for the narrow instrumental role that it assigns to education (inasmuch as HCT disregards some of important non-material aspects of education), as well as for its inability to satisfactorily reflect the cultural, gender-based, emotional and historical differences that can influence educational choices and individual well-being.

Journal articles

Ricciato, F. et al. (2019b). “Trusted smart statistics: Motivations and principles”. In: *Statistical Journal of the IAOS* 35.4, pp. 589–603. DOI: [10.3233/SJI-190584](https://doi.org/10.3233/SJI-190584). URL: <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji190584>.

In this contribution we outline the concept of Trusted Smart Statistics as the natural evolution of official statistics in the new datafied world. Traditional data sources, namely survey and administrative data, represent nowadays a valuable but small portion of the global data stock, much thereof being held in the private sector. The availability of new data sources is only one aspect of the global change that concerns official statistics. Other aspects, more subtle but not less important, include the changes in perceptions, expectations, behaviours and relations between the stakeholders. The environment around official statistics has changed: statistical offices are not any more data monopolists, but one prominent species among many others in a larger (and complex) ecosystem. What was established in the traditional world of legacy data sources (in terms of regulations, technologies, practices, etc.) is not guaranteed to be sufficient any more with new data sources. Trusted Smart Statistics is not about replacing existing sources and processes, but augmenting them with new ones. Such augmentation however will not be only incremental: the path towards Trusted Smart Statistics is not about tweaking some components of the legacy system but about building an entirely new system that will coexist with the legacy one. In this position paper we outline some key design principles for the new Trusted Smart Statistics system. Taken collectively they picture a system where the smart and trust aspects enable and reinforce each other. A system that is more extrovert towards external stakeholders (citizens, private companies, public authorities) with whom Statistical Offices will be sharing computation, control, code, logs and of course final statistics, without necessarily sharing the raw input data.

Rueda-Cantuche, J. M., A. F. Amores, and I. Rémond-Tiedrez (2019). “Can supply, use and input–output tables be converted to a different classification with aggregate information?” In: *Economic Systems Research*. DOI: [10.1080/09535314.2019.1655393](https://doi.org/10.1080/09535314.2019.1655393). URL: <https://www.tandfonline.com/doi/full/10.1080/09535314.2019.1655393>.

Every change in the product and/or industry classifications and/or methodology of supply, use and input-output tables makes any medium- to long-term policy analysis impossible unless appropriate conversions are provided by national statistical institutes using more detailed data. However, can these tables be reasonably converted to a different classification of industries and products using aggregate information? We develop a conversion method that allows changes in classification that are independent of the number of industries and products. In addition, we provide evidence about its empirical performance compared with projection methods. We find projection methods perform better than conversion methods, at least when using aggregate information. Nonetheless, unlike conversion methods, projection methods generally require supply, use and input/output tables in the new

classification that might not always be available. In their absence, we recommend using more detailed and sophisticated data.

Sanz, A. F., S. Luhmann, and A. G. Moraleda (2019). “Official Statistics through the eyes of students and teachers – The European Statistics Competition”. In: *ASTA Wirtschafts- und Sozialstatistisches Archiv* 13, pp. 245–255. DOI: [10.1007/s11943-019-00249-5](https://doi.org/10.1007/s11943-019-00249-5). URL: <https://link.springer.com/content/pdf/10.1007/s11943-019-00249-5.pdf>.

Sutcliffe, L. M. E. et al. (2019). “The LUCAS grassland module pilot – Qualitative monitoring of grassland in Europe”. In: *Palaeoartctic Grasslands* 40, pp. 27–31. DOI: [10.21570/EDGG.PG40](https://doi.org/10.21570/EDGG.PG40). URL: https://edgg.org/sites/default/files/page/Palaeoartctic_Grasslands_40_0.pdf.

The Land Use/Cover Area-Frame Survey (LUCAS) is a European inventory carried out every three years and coordinated by Eurostat. It aims to provide information for policy and science on land use, land cover and environmental parameters by surveying a statistically representative sample of points spread across the EU countries. In 2018, a new grassland module was piloted within the survey. This pilot aims to collect detailed information on the environmental and ecological quality of the grassland, as well as its type and intensity of use. Between April and July 2018, 3734 grassland points in 26 countries were surveyed using this standardised methodology. Of these points, 747 underwent an additional quality control to check the accuracy of the survey method. This is the first time a standardised methodology has been used to collect ecological data on grasslands in a coordinated manner over so wide a geographical range in Europe. The analysis of the data from this survey is ongoing, so the purpose of this article is to briefly describe the method used in the new grassland module and inform readers about how this pilot was developed.

Rueda-Cantuche, J. M., I. Rémond-Tiedrez, and M. C. Bouwmeester (2018). “Institutionalization of inter-country input-output tables: Working towards harmonization and standardization”. In: *Journal of Industrial Ecology* 22.3, pp. 485–486. DOI: [10.1111/jiec.12761](https://doi.org/10.1111/jiec.12761). URL: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/jiec.12761>.

Effective policy to encourage sustainable production and consumption is needed to shape the future so that our impact stays in line with the earth’s carrying capacity. To design and monitor effective policy, good-quality data are indispensable. Production and consumption are two sides of the same coin, and in today’s globalized world, an integrated and consistent inter-country accounting framework that links these two is a necessity for adequate analysis. A better understanding of global value chains starts with capturing production and trade relations in a coherent and complete system. More insight in the environmental impact of consumption requires an integrated environmental-economic accounting framework. Although producers generally are the ones to pay the wages, extract the resources, and emit the greenhouse gases - our productive system is in place to serve our consumer society. More awareness of the impact of consumption, at home and abroad, is needed to change our behaviour and create a sustainable economy.

Buono, D. et al. (2017). “Big data types for macroeconomic nowcasting”. In: *Eurostat Review on National Accounts and Macroeconomic Indicators (EU-*

RONA) 1, pp. 67–77. URL: <https://ec.europa.eu/eurostat/cros/system/files/euronaissue1-2017-art4.pdf>.

In this paper we present a detailed discussion on various types of big data which can be useful in macroeconomic nowcasting. In particular, we review the big data sources, availability, specific characteristics and their use in the literature. We conclude this paper identifying the big data types which could be adopted for real applications.

- Hagsten, E. and A. Sabadash (2017). “A neglected input to production: The role of ICT-schooled employees in firm performance”. In: *International Journal of Manpower* 38.3, pp. 373–391. DOI: [10.1108/IJM-05-2015-0073](https://doi.org/10.1108/IJM-05-2015-0073). URL: <https://www.emerald.com/insight/content/doi/10.1108/IJM-05-2015-0073/full/pdf?title=a-neglected-input-to-production-the-role-of-ict-schooled-employees-in-firm-performance>.

The purpose of this paper is to broaden the perspective on how information and communication technology (ICT) relates to productivity by introducing a novel ICT variable: the share of ICT-schooled employees in firms, an intangible input often neglected or difficult to measure.

- Pantea, S., F. Biagi, and A. Sabadash (2017). “Are ICT displacing workers in the short run? Evidence from seven European countries”. In: *Information Economics and Policy* 39, pp. 36–44. DOI: [10.1016/j.infoecopol.2017.03.002](https://doi.org/10.1016/j.infoecopol.2017.03.002). URL: <https://www.sciencedirect.com/science/article/pii/S0167624516301615>.

This paper examines the short run labour substitution effects of using ICT at firm-level in the manufacturing and services sectors in seven European countries, during the period 2007–2010. The data come from a unique dataset provided by the ESSLait Project on Linking Microdata, which contains internationally comparable data based on the production statistics linked at firm level with the novel ICT usage indicators. We adopt a standard conditional labour demand model and control for unobservable time-invariant firm-specific effects. The results show that ICT use has a statistically insignificant labour substitution effect and this effect is robust across countries, sectors and measures of ICT use. Our findings suggest that increased use of ICT within firms does not reduce the numbers of workers they employ.

- Rueda-Cantuche, J. M. et al. (2017). “Assessment of European use tables at basic prices and valuation matrices in the absence of official data”. In: *Economic Systems Research* 30.2, pp. 252–270. DOI: [10.1080/09535314.2017.1372370](https://doi.org/10.1080/09535314.2017.1372370). URL: <https://www.tandfonline.com/doi/full/10.1080/09535314.2017.1372370>.

Input-Output modellers are often faced with the task of estimating missing Use tables at basic prices and also valuation matrices of the individual countries. This paper examines a selection of estimation methods applied to the European context where the analysts are not in possession of superior data. The estimation methods are restricted to the use of automated methods that would require more than just the row and column sums of the tables (as in projections) but less than a combination of various conflicting information (as in compilation). The results are assessed against the official Supply, Use and Input-Output tables of Belgium, Germany, Italy, Netherlands, Finland, Austria and Slovakia by using matrix difference

metrics. The main conclusion is that using the structures of previous years usually performs better than any other approach.

Agafitei, M. et al. (2015). “Measuring output quality for multisource statistics in official statistics: Some directions”. In: *Statistical Journal of the IAOS* 31.2, pp. 203–211. DOI: [10.3233/sji-150902](https://doi.org/10.3233/sji-150902). URL: <https://content.iospress.com/download/statistical-journal-of-the-iaos/sji902?id=statistical-journal-of-the-iaos\%2Fsji902>.

Many statistical offices have been moving towards an increased use of administrative data sources for statistical purposes, both as a substitute and as a complement to survey data. Moreover, the emergence of big data constitutes a further increase in available sources. As a result, statistical output in official statistics is increasingly based on complex combinations of sources. The quality of such statistics depends on the quality of the primary sources and on the ways they are combined. This paper analyses the appropriateness of the current set of output quality measures for multiple source statistics, it explains the need for improvement and outlines directions for further work. The usual approach for measuring the quality of the statistical output is to assess quality through the measurement of the input and process quality. The paper argues that in multisource production environment this approach is not sufficient. It advocates measuring quality on the basis of the output itself - without analysing the details of the inputs and the production process - and proposes directions for further development.

Infante, E., D. Buono, and A. Buono (2015). “A 3-way ANOVA a priori test for common seasonal patterns and its application to direct versus indirect methods”. In: *Eurostat Review on National Accounts and Macroeconomic Indicators (EURONA)* 1, pp. 93–145. ISSN: 1977-978X. URL: https://ec.europa.eu/eurostat/cros/system/files/05y-newanova_techsav_dtp_final.pdf.

In this paper we propose a new a priori test to be used for the identification of a common seasonal pattern. The test is applied a priori to any running of a seasonal adjustment procedure. The test is a three way ANOVA, where the three factors are the series, the time frequency and the year. One of the possible applications of using such a test would be when selecting either the direct or indirect approach when seasonally adjusting. The Seasonally Adjusted series of an aggregate can be obtained by seasonal adjusting it (direct approach) or by aggregating the seasonally adjusted individual series (indirect approach). It should be noted that, to date, the literature has been mainly focusing on an a posteriori comparison among the results achieved by applying different approaches. This paper seeks to set out an a priori strategy for the identification of the most effective seasonal adjustment of the aggregate.

Ruggeri Cannata, R., D. Buono, and F. Biscosi (2015). “The Macroeconomic Imbalances Procedure and the scoreboard: Ensuring data coverage”. In: *Eurostat Review on National Accounts and Macroeconomic Indicators (EURONA)* 2, pp. 97–118. ISSN: 1977-978X. URL: <https://ec.europa.eu/eurostat/documents/3217494/7114363/KS-GP-15-002-EN-N.pdf>.

The Macroeconomic Imbalance Procedure (MIP) is a surveillance mechanism that aims to identify potential risks early on, prevent the emergence of harmful macroe-

conomic imbalances and correct the imbalances that are already in place. It is a system for monitoring economic policies and detecting potential harms to the proper functioning of the economy of a Member State, of the Economic and Monetary Union, and of the European Union as a whole. The MIP is supported by the analysis of a set of headline and auxiliary indicators, whose data coverage can reach twenty years due to data transformations. In order to ensure the necessary time series length for policy makers, statisticians can resort to statistical techniques such as back-calculation. This paper illustrates the MIP in the European Union policy context and some applications of back-calculation to two MIP indicators.

Conference articles

- Grazzini, J., J. Gaffuri, and J.-M. Museux (2019). “Delivering Official Statistics as Do-It-Yourself services to foster producers’ engagement with Eurostat open data”. In: *Proc. New Techniques and Technologies for Statistics (NTTS)*. DOI: [10.5281/zenodo.3240272](https://doi.org/10.5281/zenodo.3240272). URL: https://www.researchgate.net/publication/332079417_Delivering_Official_Statistics_as_Do-It-Yourself_services_to_foster_producers'_engagement_with_Eurostat_open_data.

Opening up data obviously provides the opportunity to involve actors from outside the European Statistical System - say producers, e.g., statisticians, scientists, citizens - and promote innovative, user-centric ways to tackle new and existing policy issues by co-designing statistical products. This also has the potential to increase National Statistical Offices efficiency and effectiveness. However, open data alone does not automatically translate to public participation to the decision-making. In most data provision services, there is an overarching top-down ideology since statistical processes are still owned, dictated and designed by National Statistical Offices and the final users are only involved as the receivers of the data and/or services. We propose to move away from the current approach by providing tools and software for accessing and using online data, so as to enable sharing best practices, learning from others’ experience, adopt common methodologies, enhance cooperation between data producers and data users, and further engage in Open Data-driven innovation.

- Lamboray, C. (2019). “Elementary aggregation: A not so elementary story!” In: *Meeting of the Ottawa Group*. URL: https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/elementary_aggregation_og_lamboray.pdf.

The compilation of a CPI is often presented in two stages. First, prices are aggregated without weights at the elementary level. Prices are typically obtained from dedicated surveys for which price collectors visit outlets and record the observed prices. These elementary price indices are then aggregated to the higher levels using expenditure weights. Nowadays, CPIs are becoming a multi-source statistics where prices are obtained not only from price collection in the field but also from transaction data, administrative data or from the Internet using web scraping techniques. Depending on the data source, different strategies can be adopted for constructing

the elementary aggregates and for compiling elementary price indices. A CPI may be compiled in more than two stages and weights may be available even within the elementary aggregates. With scanner data, the index compiler must make two main structural decisions which can have a significant impact on inflation measurement. First, the item which is being aggregated must be defined. Second, the level must be fixed up to which these items are first aggregated. To discuss this second issue, we distinguish two strategies for a category that can be divided into sub-categories. Either the items are directly aggregated to the category level, possibly using a multilateral method. Alternatively, the multilateral method aggregates only up to the sub-category level, and these intermediate sub-category level indices are then aggregated to the category level using for instance a Laspeyres-type index formula. The objective of this paper is to examine the impact of introducing this additional level of fixity in the CPI structure.

- Luhmann, S. et al. (2019). “Promoting reproducibility-by-design in statistical offices”. In: *Proc. New Techniques and Technologies for Statistics (NTTS)*. DOI: [10.5281/zenodo.3240198](https://doi.org/10.5281/zenodo.3240198). URL: https://www.researchgate.net/publication/332045930_Promoting_reproducibility-by-design_in_statistical_offices.

This paper emphasizes the need for Official Statistics to go beyond current practice and exceed the limits of the National Statistical Offices and the European Statistical System to reach and engage with producers - e.g. statisticians, scientists and citizens. Through the adoption of some best practices derived from the Open Source Software community and the integration of modern technological solutions, the “Shared, Transparent, Auditable, Trusted, Participative, Reproducible, and Open” principles can help create new participatory models of knowledge and information production.

- Mészáros, M. (2019). “Aggregating flags – A standardised and rational approach”. In: *Proc. New Techniques and Technologies for Statistics (NTTS)*. Online poster. URL: https://coms.events/ntts2019/data/x_abstracts/x_abstract_90.docx.

A flag is an attribute of a cell in a data set that provides additional qualitative information about the statistical value of that cell. They can indicate a wide range of information, for example, that a given value is estimated, confidential or represents a break in the time series. Currently different sets of flags are in use in the European Statistical System (ESS). Some statistical domains use the SDMX code list for observation status and confidentiality status, OECD uses a simplified version of the SDMX code lists and Eurostat uses a short list of flags for dissemination which combines the observation and confidentiality status. While in most cases it is well defined how a flag shall be assigned to an individual value, it is not straightforward to decide what flag shall be propagated to aggregated values like a sum, an average, quantiles, etc. This topic is important for Eurostat as the European aggregates are derived from national data points. Thus the information contained in the individual flags need to be summarized in a flag for the aggregate. This issue is not unique to Eurostat, but can occur for any aggregated data. For example, a national statistical institute may derive the national aggregate from regional data sets. In addition, the dissemination process provides further peculiarity: only a

limited set of flags, compared to the set of flags used in the production process, can be applied in order to make it easily understandable to the users. In the scientific community there is a wide range of research about the consequences of data aggregation but it concentrates only on the information loss during aggregation of information and there is no scientific guidance how to aggregate flags. This paper is an attempt to provide a picture about the current situation and provide some systematic guidance how to aggregate flags in a coherent way. Eurostat is testing various approaches with a view to well balance transparency and clarity of the information made available to users in a flag. From several options, 3 methods (hierarchical, frequency and weighted frequency) are implemented in an R package for assigning a flag to an aggregate based on the underlying flags and values. Since the topic has relevance outside of Eurostat as well, it was decided to publish the respective code with documentation with a view to foster re-use within the European Statistical System and to stimulate discussion, including with the user community.

Rémond-Tiedrez, I. and J. M. Valderas Jaramillo (2019). “The Eurostat’s balanced view of trade in services”. In: *Proc. International Input-Output Conference*. URL: <https://www.iioa.org/conferences/27th/papers/files/3736.pdf>.

Asymmetries due to the mismatching in the data provided by one country and the mirror flow provided by its partner country for the same transactions are an important issue in trade statistics, especially when it comes to link all European Union (EU) economies as in the FIGARO dataset. Although at EU level, balance of payments statisticians and trade in services statisticians follow up regularly on the asymmetries and try to reduce them, we needed to implement a methodology for compiling a balanced view of trade in services as an input to the EU inter-country supply, use and input-output tables. For the first release of FIGARO tables for the year 2010, the 2010 international trade in services data (ITSS) serves as the primary input. Their exports and imports (or mirror exports) are subsequently cleaned, imputed, estimated, modelled and confronted with Balance of Payments data to get a full dataset for 29 countries (EU Member States plus USA), 30 partner countries (plus RoW) and a number of services items. The balancing of the resulting exports and import values to solve the bilateral trade asymmetries is based on the methodology developed by the European Commission and the OECD. As the EU inter-country supply, use and input-output tables present economies using the activity and product classification, the last step is to bridge the balanced trade view of the data from services categories to product classification (CPA/CPC). The paper summarises the steps as compiled for the 2010 tables and shows the way foreseen to improve the compilation steps for the time series 2010-2016. We also evaluate the impact on the original input data of each of the steps involved.

Ricciato, F. and A. Bujnowska (Mar. 2019). “Privacy and data confidentiality for Official Statistics: New challenges and new tools”. In: *Proc. New Techniques and Technologies for Statistics (NTTS)*. URL: https://coms.events/ntts2019/data/x_abstracts/x_abstract_190.pdf.

The modern society is undergoing a process of massive datafication. The availability of new digital data sources represents an opportunity for Statistical Offices (SO)

to complement traditional statistics as well as to produce novel statistical products with improved timeliness and relevance. However, such opportunities come with important challenges in almost every aspect – methodological, business models, data governance, regulation, organizational and others. The new scenario calls for an evolution of the *modus operandi* adopted by SO also with respect to privacy and data confidentiality, that is the focus of the present contribution. We propose here a discussion framework focused on the prospective combination of advanced Statistical Disclosure Control (SDC) methods with Secure Multi-Party Computation (SMC) techniques.

Ricciato, F., G. Lanzieri, and A. Wirthmann (2019). “Towards a methodological framework for estimating present population density from Mobile Network Operator data”. In: *Proc. workshop on the use of Administrative Data and Social Statistics*. URL: https://ec.europa.eu/eurostat/cros/system/files/mno_spatial_density_ricciato_lanzieri_wirthmann_2019_v1.pdf.

The concept of ‘present population’ is gaining increasing attention in official statistics. The (almost) continuous measurement of present population provides a basis to derive indicators of population exposure that are relevant in different application domains. One possible approach to measure present population exploits data from Mobile Network Operators (MNO), including CDR but also more informative (and complex) signalling records. Such data, collected primarily for network operation processes, can be repurposed to infer patterns of human mobility. Two decades of research literature have produced several case studies (mostly limited to CDR data) and a variety of ad-hoc methodologies tailored to specific datasets. Moving beyond the stage of explorative research, towards production of official statistics, requires a more systematic and sustainable approach to methodological development. Towards this aim, Eurostat and other members of the European Statistical System (ESS) are working towards the definition of a general Reference Methodological Framework. In this contribution we report on the methodological aspects related to the estimation of present population density, for which we present a general and modular methodological structure. Along the way, we identify a number of specific research (sub)problems requiring further attention by the research community.

Ricciato, F. and A. Wirthmann (2019). “Trusted Smart Statistics: How new data will change Official Statistics”. In: *Proc. Data for Policy conference*. DOI: [10.5281/zenodo.3066060](https://doi.org/10.5281/zenodo.3066060). URL: https://zenodo.org/record/3066061/files/ricciato_wirthmann_Data4Policy_2019.pdf.

In this discussion paper we outline the motivations and the main principles of the Trusted Smart Statistics concept under development in the European Statistical System (ESS) to respond to the challenges posed by the prospective use of innovative digital data sources for the production of official statistics.

Ricciato, F., A. Wirthmann, and M. Hahn (June 2019). “Integrating alternative data sources into Official Statistics: A system-design approach”. In: *Proc. Conference of European Statisticians (CES)*. United Nations Economic Commission for Europe. URL: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2019/ECE_CES_2019_32-Eurostat.pdf.

New types of digital data sources (or "big data") are now available as by-product of other technological processes. New data sources differ from the traditional data sources in use for official statistics, namely survey data and administrative records, along multiple dimensions. Therefore, the adoption of new data sources for the regular production of official statistics requires innovations at multiple levels, including new processing paradigms, computation methods, data access and governance models, staff skills, etc. The term "Trusted Smart Statistics" was put forward by Eurostat to indicate a comprehensive framework to evolve official statistics towards adoption of new data sources along with traditional ones. In this document we focus on the need to take a systemic view, and in general a system-design approach, towards the development of novel processing methodologies for new types of data. We argue for the need to identify selected 'classes' of new data types (e.g., mobile network operator data, smart energy meters, satellite images, etc.) and, for each class, to build a general Reference Methodological Framework as basis for developing specific methodologies for particular use-cases and statistical products. We discuss the principles that should inform the construction of such framework, and briefly report on the ongoing work being conducted at Eurostat for one particular class of data, namely mobile network operator data.

Ricciato, F. et al. (2019a). "A reflection on privacy and data confidentiality in Official Statistics". In: *Proc. International Statistical Institute (ISI) World Statistics Congress*. URL: https://www.bis.org/ifc/events/isi_wsc_62/ips177_paper3.pdf.

The availability of new digital data sources represents an opportunity for Statistical Offices (SO) to complement traditional statistics and/or deliver novel statistics with improved timeliness and relevance. Nowadays SOs are part of a larger "data ecosystem" where different organizations, including public institutions and private companies, engage in the collection and processing of different kinds of (new) data about citizens, companies, goods etc. In this multi-actors scenario it is often desirable to let one organization extract some output statistics (i.e., aggregate information) from input data that are held by other organization(s) in different administrative domain(s). We refer to this problem as cross-domain statistical processing. To achieve this goal, the most intuitive approach - but not the only one - is to exchange raw input data across administrative domains (organizations). However, this strategy is not always viable when personal input data are involved, due to a combination of regulatory constraints (including lack of explicit legal basis for data sharing), business confidentiality, privacy requirements, or a combination of the above. Furthermore, new data sources often embed a much more pervasive view about individuals than traditional survey and/or administrative data, an aspect that amplifies the potential risks of data concentration. In such cases, performing cross-domain statistical processing requires technologies to elicit only the agreed-upon output information (exactly or approximately) without revealing the input data. This entails addressing two distinct but complementary problems. First, we need to compute the desired output statistics without seeing the raw input data. Second, we need to control the amount of information that might be inferred about individual data subjects in the input dataset from the output. In the field of privacy engineering the notions of "input privacy" and "output pri-

vacy” are used to refer respectively to these two problems. We remark that these problems are separable, i.e., they can be addressed with distinct tools and methods that get combined together, overlaid or juxtaposed. In this contribution we review recent advances in both fields and briefly discuss their complementary roles. As for input privacy, we provide a brief introduction to the fundamental principles of Secure Multi-Party Computation (SMPC). As for output privacy, we review recent advances in the field of Statistical Disclosure Control (SDC). Finally, we discuss possible scenarios for SMPC and SDC integration in the future ”confidentiality engineering” setup of modern official statistics.

Rueda-Cantuche, J. M., A. Velazquez-Afonso, and I. Rémond-Tiedrez (2019).

“Traceability of the assumptions made in the construction of the EU inter-country supply, use and input-output tables”. In: *Proc. International Input-Output Conference*. Online abstract. URL: https://www.iioa.org/conferences/27th/papers/files/3855_20190423101_FIGARO_book_chapter13.pdf.

The modular approach adopted in the construction of the EU inter-country supply, use and input-output tables (Figaro project) to map the different adjustments and imputations made to the original data allows each adjustment/imputation to be measured at the different stages of the compilation process. As a result, this paper provides summary statistics based on: the comparison between the international merchandise and services trade data adjusted for goods sent abroad for processing and merchanting activities and the trade values in the national available SUTs (i.e. discrepancies); the analysis of the row and column total discrepancies by countries, users and products; the analysis of the final balancing adjustments made to estimate the inter-country use table without discrepancies, by countries, users and products. This analysis provides useful information for the user of the FIGARO tables and helps producers in: highlighting the importance of the scope of some of the statistics they produce; identifying what type of data is still missing from national statistical offices; and identifying where to put more efforts in future revisions. All these aspects are relevant for the compilation process.

Selenius, J. et al. (2019). “Agricultural census 2020 – How to reduce costs and burden? The European Statistical System approach”. In: *Proc. International Statistical Institute (ISI) World Statistics Congress*.

Wirtz, C., J. Selenius, and A. C. Lazar (2019). “Modernisation of the European Agricultural Statistics System (EASS): Strategy for agricultural statistics 2020 and beyond”. In: *Proc. International Statistical Institute (ISI) World Statistics Congress*.

Bach, F., W. Kloeck, and A. Bujnowska (2018). “Statistical confidentiality: New initiatives in the European Statistical System”. In: *Proc. Quality conference*. Online presentation. URL: https://www.q2018.pl/wp-content/uploads/Sessions/Session%2031/Fabian%20Bach/Session%2031_%20Fabian%20Bach.docx.

The protection of confidential information has a huge impact on how statistical data can be published and used for analysis, which makes it a key aspect of data quality. This paper presents new methods and tools currently being investigated in the ESS in order to publish more - and more useful - data without compromising statistical confidentiality. It covers new methodological and IT developments,

where concrete use cases demonstrate their impact on data quality. For instance, a promising methodological direction is random noise: several ESS use cases at different maturity stages are presented, including recommendations for the harmonised protection of 2021 EU Census data. Another direction is to reflect at a more fundamental level where protection is needed. Several ideas will be presented along this line.

Florescu, D. C. (2018). “European structural farm statistics – New quality rating system”. In: *Proc. Quality conference*. Online presentation. URL: https://www.q2018.pl/wp-content/uploads/Sessions/Session\%2019/Denisa\%20Florescu/Session\%2019_Denisa\%20Florescu.DOCX.

Eurostat, together with the statistical bodies belonging to the European Statistical System (ESS), has adopted a quality rating system to guide the dissemination of structural farm statistics derived from farm structure surveys. The system does this by showing when the estimates are sufficiently reliable to be published, either with or without a warning. It is based on: (i) coefficients of variations for totals and means of continuous variables; (ii) standard errors for proportions and counts. This paper also presents the work carried out to harmonise variance estimation methods and their application within the ESS. To apply the new quality rating system consistently, Eurostat and the national statistical bodies must compute roughly the same variance estimates. Future structural farm statistics will come from the data collected from ‘Integrated Farm Statistics’, based on a modular approach. This will lead to more complex national sampling designs. The paper also outlines ongoing developments towards integrating additional sampling design information specific to national multi-stage sampling in the estimation of variance. The paper also introduces new quality reporting based on the European Standard Quality Reporting System (ESQRS) template. This is of great help in assessing all quality dimensions, thereby improving the quality of EU data and metadata. Farm structure surveys are the main source of information on the current state of agriculture and the trends it is undergoing, required to monitor the common agricultural policy and other EU policies. High-quality data are essential for decision-makers.

Grazzini, J., J.-M. Museux, and M. Hahn (2018). “Empowering and interacting with statistical producers: A practical example with Eurostat data as a service”. In: *Proc. Conference of European Statistics Stakeholders (CESS)*. DOI: [10.5281/zenodo.3240557](https://doi.org/10.5281/zenodo.3240557). URL: https://www.researchgate.net/publication/325973362_Empowering_and_interacting_with_statistical_producers_a_practical_example_with_Eurostat_data_as_a_service.

While the importance of openness and transparency in statistical processes, and how these can be supported through open algorithms and open data, has been already emphasized, this contribution aims at showcasing an approach where algorithms and data are delivered as interactive, reusable and reproducible computing services. This will eventually provide producers with the necessary tools to perform, for themselves, data analytics on Eurostat data in a straightforward manner.

Grazzini, J. et al. (2018). ““Show me your code, and then I will trust your figures”: Towards software-agnostic open algorithms in statistical production”. In: *Proc. Quality conference*. DOI: [10.5281/zenodo.3240282](https://doi.org/10.5281/zenodo.3240282). URL:

https://www.researchgate.net/publication/325320551_Show_me_your_code_and_then_I_will_trust_your_figures_Towards_software-agnostic_open_algorithms_in_statistical_production.

This contribution aims at further promoting the development and deployment of open, reproducible, reusable, verifiable, and collaborative computational resources in statistical offices regardless of the platform/software in use. Motivated by the consensus that data-driven evidence-based policymaking should be transparent, we argue that such approach is not only necessary for the practical implementation of statistical production systems, but also essential to reinforce the quality and trust of official statistics, especially in the context of a 'post-truth' society. We discuss some practical requirements to gear the continuous and flexible development and deployment of software components in production environments. Together with the adoption of some best practices derived from the open source community and the integration of new technological solutions, we propose to unleash the social power of open algorithms so as to create new participatory models of interaction between producers that can contribute to a more holistic and extensive approach to production systems. Overall, a greater transparency in designing production processes is expected to result in a better grip on the quality of the statistical processes involved in data-driven policy-making. We illustrate this flexible and agile approach with various open, stand-alone software or source code used in statistical production environments at Eurostat.

Martins Ferreira, P., I. Rémond-Tiedrez, and J. M. Rueda-Cantuche (2018). "QDR methodology: Understanding bilateral trade flows in the European Union". In: *Proc. International Input-Output Conference*. [Online abstract](https://www.iioa.org/conferences/26th/papers/files/3348_20180515021_iioa2018_QDR.pdf). URL: https://www.iioa.org/conferences/26th/papers/files/3348_20180515021_iioa2018_QDR.pdf.

Trade asymmetry has been a well-known fact and there are extensive literature and reports about the causes for those asymmetries. There is also a recognised effort made by trade statisticians for mitigate trade asymmetry over time. Notwithstanding the positive achievements that have been made so far, to build an Inter-Country Supply, Use and Input-Output tables (IC-SUIOT) we more than low trade asymmetry: we need no trade asymmetry at all. The European Statistical System (ESS) has an extensive and rich amount of trade data and a lot of resources are devoted to measure trade flows. Nevertheless, the customs union of the EU adds another challenge regarding trade in goods statistics: Member-States declare imports/exports for customs or tax purposes without this Member State having acquired ownership of the goods, i.e. declare quasi-transit as well. While relevant for physical flow of trade, quasi-transit and re-exports distort the geographical economic relationship among Member-States and therefore they should be identified and taken into account in the framework of IC-SUIOT. QDR methodology was developed in order to address the specificities of trade in goods in EU by providing a way to estimate consolidated trade flows, i.e. solving trade asymmetries, between two countries by three types of trade: quasi-transit (Q), domestic (D) and re-export (R). For quasi-transit and re-exports the intermediary country between that takes part of the physical flow between origin and destination is also identified. QDR methodology

was used in FIGARO project and it revealed very useful for identifying relevant trade relationships within countries.

Ricciato, F. (2018). “Towards a reference methodological framework for processing MNO data for Official Statistics”. In: *Proc. Global Forum on Tourism Statistics*. URL: http://www.15th-tourism-stats-forum.com/pdf/Papers/S3/3_1_A_Reference_Methodological_Framework_for_processing_mobile_network_operatordata_for_official_statistics.pdf.

Mobile network signalling data, captured from the continuous interaction of mobile terminals with the cellular network, have better spatial/temporal resolution than traditional Call Detail records (CDR). However, their format and semantic are intimately connected with network-specific technical aspects. For this reason, such data are considerably more complex and have a higher degree of heterogeneity across different Mobile Network Operators (MNO). It is difficult for experts outside the telecommunication domains, such as e.g. statisticians, to interpret and manipulate such data directly. In the proposed contribution we present a general Reference Methodological Framework (RMF) intended to facilitate the use of signalling data by statisticians. The RMF is inspired by the principles of functional layering and by the “hour-glass model”, which lie at the foundation of modern computer network architectures. The RMF encompasses a convergence layer that decouples the complexity of signalling data at the bottom from the statistical definitions on the top. This allows experts from the two domains, MNO engineers and statisticians, to work independently and eases the evolution of the two layers. This paper presents the general principles underlying the RMF, the role and responsibilities of the different actors in transforming elemental data into meaningful and relevant statistical concepts, provides a concrete actionable proposal and presents early results from its application in a pilot project conducted in collaboration between Eurostat and one European MNO. We highlight lessons learned and give an outlook for the future development and implementation of the RMF and its application to tourism statistics and other areas of statistics.

Ricciato, F. et al. (2018a). “Processing of Mobile Network Operator data for Official Statistics: The case for public-private partnerships”. In: *Proc. Conference of the Directors General of the National Statistical Institutes (DGINS)*. Online presentation. URL: http://www.dgins2018.ro/wp-content/uploads/2018/10/17-MNO-data-for-Official-Statistics-DGINS_v35b_final.pdf.

This paper discusses various aspects related to the potential partnership between Statistical Offices (SO) and Mobile Network Operators (MNO) to leverage MNO data for the computation of official statistics. MNO data are complementary to other data sources that are already available to SOs (e.g., survey data, administrative registers) and their combination can lead to a new generation of statistical products, delivered more timely and with better spatio-temporal resolution than traditional statistics. This enables statisticians to gain more accurate and up-to-date insight into various aspects of human mobility and related socio-economic phenomena (e.g., tourism flows, presence and residence, commuting patterns, use of transportation means among others) with clear advantages for the process of policy design and evaluation based on such statistics. The cooperation between SO

and MNO can be designed to prevent potential conflicts between the public and private interests, e.g. by the provision of adequate protection for business confidentiality, methodological quality and process transparency. We argue that partnering with SO brings direct and indirect benefits also to the MNOs, particularly in terms of empowering the portfolio of commercial analytic products they can offer to business customers. Synergies between the production of official statistics and commercial analytic products can be positively leveraged within the framework of a well-designed partnership model. By doing so, the SO-MNO partnership does not represent as a risk to the MNO business nor a diminution of the role and independency of SO, but rather as an additional opportunity for both sides. While the focus of this paper is on partnership models between SOs and MNOs, many elements of the discussion apply as well to private data holders from other sectors, and may contribute to advance the future vision of public-private partnerships for joint data analytics.

Ricciato, F. et al. (2018b). “Towards a reference architecture for Trusted Smart Statistics”. In: *Proc. Conference of the Directors General of the National Statistical Institutes (DGINS)*. URL: https://www.researchgate.net/publication/328215827_Towards_a_Reference_Architecture_for_Trusted_Smart_Statistics.

In this contribution we outline the concept of Trusted Smart Statistics as the natural evolution of official statistics in the new datafied world, where traditional data sources (survey and administrative data) represent a valuable but small portion of the global data stock, much thereof being held in the private sector. In order to move towards practical implementation of this vision a Reference Architecture for Trusted Smart Statistics is required, i.e., a coherent system of technical, organisational and legal means combined to provide an articulated set of trust guarantees to all involved players. In this paper we take a first step in this direction by proposing selected design principles and system components that, as of the current state of play, we believe will be part of the final design. The goal of this contribution is not to propose a ready-made fully-fledged solution, but rather build awareness about the necessary elements (technological and not) and fuel the discussion with the relevant stakeholders.

Rueda-Cantuche, J. M. et al. (2018a). “Employment effects of EU services exports to the rest of the world by modes of supply using the Eurostat’s EU inter-country input-output tables”. In: *Proc. International Input-Output Conference*. URL: <https://www.iioa.org/conferences/26th/papers/files/3345.pdf>.

Services are increasingly delivered across borders under various modes of supply and gaining higher shares over all the economic activities. However, the availability of statistics on the international supply of services detailed by services category, mode of supply and partner country is limited and at the same time critically important for trade policy making. Based on the most recent Eurostat published data, this paper presents the first attempt to estimate the employment effects by modes of supply using official statistics and the Eurostat’s experimental EU Inter-country Input-Output Table (FIGARO Project).

Rueda-Cantuche, J. M. et al. (2018b). “From theory to practice: What makes

the European Union’s inter-country supply, use and input-output tables different?” In: *Proc. International Input-Output Conference*. Online abstract. URL: https://www.iioa.org/conferences/26th/papers/files/3338_20180515031_iioa2018_FIGARO_main.pdf.

The Eurostat-JRC project ”Full International and Global Accounts for Research in Input-Output Analysis” (FIGARO) has produced experimental EU-Inter Country Supply, Use and Input-Output Tables for the year 2010 in line with the ESA 2010 methodology. Setting up a European Inter-country Supply, Use and Input-Output Table implies the compilation of a balanced view of international trade consistent with National Accounts data. It is therefore absolutely necessary to: (a) reconcile the trade asymmetries and provide one single trade flow for each bilateral transaction between partners; and (b) align the trade figures with National Accounts data, in order to capture, for instance, the potential environmental, social and economic effects of supply and demand shocks on the national economies via the existing global value (and supply) chains. The paper describes methodological issues raised by the construction process of the Inter-country Supply, Use and Input-Output Tables: e.g. econometric estimations of cif/fob margins; econometric estimations of missing bilateral services trade; alignment of trade statistics and national accounts data: e.g. goods sent abroad for processing, merchanting activities.

Salvati, M. and M. Mészáros (2018). ”Introduction to ”flagr””. In: *Proc. conference on use of R in Official Statistics (uRos)*. URL: <http://r-project.ro/conference2018/uRos2018.pdf#page=54>.

The object of this paper is to present the R package ’flagr’ that is in development in Eurostat for facilitating the internal revision of the use of flags and flagging of aggregates in dissemination. The ’flagr’ package provides general functions following the methodological guidelines suggested by the SDMX for the aggregate. The package provides three different functions how the individual flags can be transferred to the aggregate. The first one is the hierarchy of the SDMX flags suggested by the implementation guidelines. This method compares all flags of a given dataset and keeps the flag for the aggregate with the highest score on the SDMX hierarchy or in a personally specified order. The second method counts the occurrences of the flags in the underlying data and the flag for the aggregate will be the flag that has the highest count. The last method not only counts the frequency of a flag is represented in the dataset, but also it takes into account the weight of the individual values, as the contribution of the corresponding individual value to the aggregate. The flag, which has the highest summed weight, is used for the flag of the aggregate if it is above a certain threshold.

Vāju, S. C. and Mészáros M. (2018). ”Administrative data and quality – Guidelines towards better quality of administrative data”. In: *Proc. Quality conference*. Online presentation. URL: https://www.q2018.pl/wp-content/uploads/Sessions/Session\%2037/M\%C3%A1ty\%C3%A1s\%20M\%C3%A9sz\%C3%A1ros/Session\%2037_Matyas\%20Meszaros.docx.

Statistical authorities need to produce data faster in a cost effective way, to become more responsive to users’ demands, while at the same time providing high quality output. One way to fulfil this is to make more use of already available data sources, and in particular administrative sources, most typically used in combination with

other sources. Depending on the use of the administrative sources and the data configuration different statistical tasks must be applied. Usually it is not only one task but a sequence of different tasks that have to be applied, for example, data integration, imputation and editing or tabulation. For these tasks different methods are available and depending on the input data quality and the data configuration the same method can have limited use or produce lower quality outputs. The use of administrative data sources risks impacting negatively quality on several dimensions, in particular accuracy and comparability. Surveys and administrative sources have both particular strengths and weaknesses. Combining them may overcome these weaknesses, provided that suitable methodology and tools are used. At the same time, harmonised measures of quality for outputs that combine administrative sources with other sources (surveys) are necessary to ensure that European Union official statistics are of sufficient quality and fit for their intended use. This paper looks at the most frequent methodological challenges faced when integrating administrative sources and provides, for typical situations, preferred methods to have the best quality of statistical output. It also introduces the work of ESSnet on the Quality of Multisource Statistics (KOMUSO) to develop quality measures and guidelines related to the use of administrative sources.

Velazquez-Afonso, A. et al. (2018). “Making the circle square: treatment of goods sent abroad for processing in the construction of the European Union’s inter-country supply, use and input-output tables”. In: *Proc. International Input-Output Conference*. Online abstract. URL: https://www.iioa.org/conferences/26th/papers/files/3347_20180515071_iioa2018_FIGARO_GSA.pdf.

The extension from national to inter-country Supply, Use and Input-Output tables (SUIOTs) consists in splitting national SUTs domestic exports (FOB) by country of destination (and importing industry) and by type of use (intermediate or final), which in turn produces indirect estimations of imports of intermediate and final goods and services among countries of origin (and exported products). It could also be the other way round, splitting national SUTs imports by countries of origin, as in the WIOD approach. The two approaches should not differ, in principle, as long as the view of bilateral trade among countries is balanced at the level of each good and service and both exports and imports are valued in FOB. However, this is not the case in official statistics, mostly due to trade asymmetries and the different valuation of exports (FOB) and imports (CIF). This paper however justifies the first choice for various reasons and put a special focus on the treatment of goods sent abroad for processing, including some indications about the necessary assumptions made in the absence of official data about trading partners and type and destination of the processed goods.

Bujnowska, A. (2017). “Statistical confidentiality in European business statistics”. In: *Proc. work session on Statistical Data Confidentiality*. United Nations Economic Commission for Europe/Eurostat. URL: https://www.unecce.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/1_confidentiality_europe.pdf.

Capaccioli, M. (2017). “The Eurostat Process Management Framework”. In: *Proc. workshop on Implementing Efficiencies and Quality of Output*. United

Nations Economic Commission for Europe. URL: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2017/mtg4/Paper_5_-_PMF_Eurostat.pdf.

The statistical organisations are facing several challenges: their mission is evolving, the budget and human resources are shrinking and new IT technologies are appearing on the market. They need to improve their rapidity to respond to new user requirements and maintain at the same time high quality products and services. In this context, managing business processes is an important sign of maturity and efficiency in organisations. Eurostat has decided to launch the project Process Management Framework (PMF) with the objective to build a harmonised documentation of the Eurostat processes, increase the process management maturity and create a pool of competence for business process modelling. This project is strongly linked with the Quality review initiative undertaken by Eurostat.

Grazzini, J. and P. Lamarche (2017). “Production of social statistics... goes social!” In: *Proc. New Techniques and Technologies for Statistics (NTTS)*. DOI: [10.5281/zenodo.3240501](https://doi.org/10.5281/zenodo.3240501). URL: https://www.researchgate.net/publication/324208747_Production_of_social_statistics_goes_social.

The scope of this paper is to present a practical framework adopted for the integration of software applications into a statistical production chain. The focus is the actual implementation of a high-level collaborative platform aiming not only at producing social statistics, but also at further fostering experimentation and analysis in that field. In doing so, we strongly support the (obvious) claim that “the modernisation and industrialisation of official statistical production needs a unified combination of statistics and computer science in its very principles”. Motivated by the consensus that processes - in particular statistical processes - for data-driven policy should be transparent, we naturally promote open, reproducible, reusable, verifiable, and collaborative software development and deployment in a statistical organisation. Beyond just devising guidelines and best practices, we show how the platform is implemented for the production of social statistics. For that purpose, we adopt a reasonable mix of bottom-up (from low-level scope to high-level vision) and top-down (from black-box process models to traceable functional modules) designs, so as to “think global, [and] act local”. In building the parts while planning the whole, we provide with a flexible and agile approach to immediate needs and current legacy issues, as well as long-term problems and potential future requirements for statistical production.

Lazar, A. C., J. Selenius, and M. Jortay (2016). “Strategy for agricultural statistics 2020 and beyond: for the future European Agricultural Statistics System (EASS).” In: *Proc. International Conference on Agricultural Statistics*. DOI: [10.1481/icasVII.2016.f37c](https://doi.org/10.1481/icasVII.2016.f37c). URL: <https://www.istat.it/storage/icas2016/f37-lazar.pdf>.

Many important policies of the European Union, such as the Common Agricultural Policy, depend on agricultural statistics. These statistics need to be of high quality, coherent, comparable and flexible, and should be produced efficiently based on users’ needs in order to best serve evidence-based policy making and monitoring. The current EU agricultural statistics system does not fulfil these requirements

well enough. To address this, Eurostat launched the "New legislation on Agricultural Statistics for a strategy towards 2020 and beyond" initiative in 2014. It aims to introduce two new legal frameworks stepwise: an "Integrated Farm Statistics" Regulation which will provide the basis for collecting farm level micro-data, based on a modular approach with core, module and ad hoc surveys; and a "Statistics on Agricultural Input/Output" Regulation which will provide aggregated statistics in tabular form. These frameworks will contain basic elements such as scope, precision and quality requirements and will use common definitions and classifications, while more technical elements will be covered by secondary legislation. EU Member States will be free to choose data sources, including administrative and other new data sources. This paper presents the Strategy for agricultural statistics 2020 and beyond and shows its suitability to meet technical and methodological requirements as well as to successfully navigate the complex institutional, legal and political context within the European Union and its 28 Member States. It can therefore serve as an instructive example for a cross-border implementation of the United Nations Global Strategy to improve agricultural and rural Statistics.

Rémond-Tiedrez, I., A. F. Amores, and J. M. Rueda-Cantuche (2016). "Development of a quality adjusted labour productivity index in the European Union – Example of the employment embodied in European exports". In: *Proc. International Input-Output Conference*. URL: <https://www.iioa.org/conferences/24th/papers/files/2341.pdf>.

The paper will introduce the methodology for the Quality Adjusted Labour Index (QALI) in the European Union which combines macro-data from National Accounts (which are the benchmarked data) and micro-data from the EU statistics of the Labour Force survey (LFS) and the Structural Earnings Survey (SES). The Quality Adjusted Labour Input is constructed for the EU-28, EA-19 and each EU MS, whenever data are available, for the full time series from 2002 to 2013, with possible extension to 2014. Survey-based data of hours worked and earnings for 2002-2007 are converted from NACE Rev.1.1 to NACE Rev.2. The QALI values by EU Member State are weighted by skills, by age and by combinations of skill and age groups. The industry breakdown varies depending on countries due to reliability/confidentiality constraints of the survey data: 21 industries (A21) for some countries, EU28 and EA19; 10 industries (A10); and the total economy. Connected to the decomposition of the volume by type of workers (by age and by skill), the results will give interesting insights on what kind of employment is supported by European exports in terms of age, qualifications, and in which industrial activities. The results will be based on the European consolidated Supply, Use and Input-Output Tables produced annually by Eurostat.

Infante, E. and D. Buono (2013). "New technique for predictability, uncertainty, implied volatility and statistical analysis of market risk using SARIMA forecasts intervals". In: *Proc. New Techniques and Technologies for Statistics (NTTS)*. URL: https://ec.europa.eu/eurostat/cros/system/files/NTTS2013fullPaper_143.pdf.

Market price data plays an essential role when aiming at the production of quality statistics to be used by policy makers at EU level. Market risk can be defined as the risk of losses in positions arising from movements in market prices, generally linked

to the risk that commodity prices and/or their implied volatility will change. From the analyst's perspective, the most informative data are possibly located within the end-series observations. This paper proposes a new Technique for Statistics to be used for assessing the presence of commodity risk that might cause market risk. The underlining idea is to assess whether the realized prices for a determined product in a specified time span is significantly apart from the SARIMA forecasts intervals. Such procedure aims also at identifying the type of eventual outliers present within the end-series observations. An applied case study on agricultural price statistics in Italy is here presented.

Infante, E., D. Buono, and A. Buono (2013). "IB test for direct versus indirect approach in seasonal adjustment". In: *Proc. New Techniques and Technologies for Statistics (NTTS)*. URL: https://ec.europa.eu/eurostat/cros/system/files/NTTS2013fullPaper_143.pdf.

The seasonally adjusted series of an aggregate can be obtained by seasonal adjusting it ("direct approach") or by aggregating the seasonally adjusted individual series ("indirect approach"). The literature to date has mainly focused upon an a posteriori comparison among the results achieved by applying different approaches. Here a new a priori test (IB test) for choosing between direct and indirect approach in seasonal adjustment is proposed. The test is applied before running any seasonal adjustment procedure. When the individual series present common seasonal patterns the aggregate will be adjusted directly, otherwise an indirect approach could be preferred. Sections 3 and 4 include a simulation and a case study, respectively. This paper seeks to set out an a priori strategy for the identification of the most effective seasonal adjustment approach to be used.

Reports and manuals

Buono, D. et al. (2018). *ESS guidelines on temporal disaggregation, benchmarking and reconciliation*. DOI: [10.2785/846595](https://doi.org/10.2785/846595). URL: <https://ec.europa.eu/eurostat/documents/3859598/9441376/KS-06-18-355-EN.pdf>.

In official statistics there is an increasing demand for indicators at a higher frequency than those that have traditionally been observed. Direct measures of indicators at a high frequency can be very costly and difficult to achieve sometimes resulting in low quality results when the information set is not adequate. In such situations temporal disaggregation techniques can constitute a feasible alternative to the direct estimation of high frequency indicators. Additionally, even when high frequency indicators can be directly compiled, they are often not consistent over time with lower frequency versions. For example, annual surveys with larger samples may give more accurate estimates of the level of a variable compared to estimates from a small monthly survey that is designed to provide estimates of monthly change. Under the hypothesis that low frequency indicators are more reliable than high frequency ones, benchmarking techniques can be used to ensure the time consistency between high and low frequency indicators. Finally, directly or indirectly measured high frequency indicators may not necessarily meet required accounting and aggregation constraints. If that low frequency indicators meet accounting

and aggregation constraints, reconciliation techniques can be used to restore them on high frequency indicators too. Eurostat and the European Statistical System (ESS) developed these guidelines to help data producers derive high frequency data (e.g. quarterly or monthly) from low frequency data (e.g. annual) and to address related temporal and accounting constraints. Typical applications are known as temporal disaggregation, benchmarking, and reconciliation. The guidelines identify best practice to: (i) achieve harmonization across national processes; (ii) enhance comparability between results; (iii) ensure consistency across domains and between aggregates and their components. The establishment of common guidelines for temporal disaggregation within the European Statistical System (ESS) is an essential step towards better harmonization and comparability of official statistics, especially in macroeconomic indicators and labour market statistics. These guidelines address the need for harmonization expressed by users from European and National Institutions. This document presents both theoretical aspects and practical implementation issues in a user friendly and easy to read framework. They meet the requirement of principle 7 (Sound Methodology) of the European Statistics Code of Practice (CoP), and their implementation is consistent with principles 14 (Coherence and Comparability) and 15 (Accessibility and Clarity). The guidelines also provide transparency of temporal disaggregation, benchmarking and reconciliation practices by encouraging documentation and dissemination of practices. These guidelines are complementary to the ESS guidelines on seasonal adjustment (Eurostat, 2015 edition), the ESS guidelines on revision policy for PEEIs (Eurostat, 2013 edition), and the Eurostat and United Nations handbook on rapid estimates. They are also in line with the Handbook on quarterly national accounts (Eurostat, 2013 edition).

Capaccioli, M., L. Gramaglia, and M. Pellegrino (2018). *Validation and Transformation Language user and reference Manual*. Statistical Data and Metadata eXchange (SDMX). URL: <https://sdmx.org/wp-content/uploads/VTL-2.0-package-2018.07.12.zip>.

Marcellino, M. G. et al. (2018). *Big data econometrics: Now casting and early estimates*. Tech. rep. 82. Research Paper Series. BAFFI-CAREFIN Centre. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3206554.

This paper aims at providing a primer on the use of big data in macroeconomic nowcasting and early estimation. We discuss: (i) a typology of big data characteristics relevant for macroeconomic nowcasting and early estimates, (ii) methods for features extraction from unstructured big data to usable time series, (iii) econometric methods that could be used for nowcasting with big data, (iv) some empirical nowcasting results for key target variables for four EU countries, and (v) ways to evaluate nowcasts and ash estimates. We conclude by providing a set of recommendations to assess the pros and cons of the use of big data in a specic empirical nowcasting context.

Buono, D., A. F. Amores, and I. Rémond-Tiedrez (2017). *Data analytics: European wheel of competitiveness*. Tech. rep. Statistical Working Papers. Eurostat. DOI: [10.2785/550234](https://doi.org/10.2785/550234). URL: <https://op.europa.eu/en/publication->

[detail/-/publication/5ce64720-41ed-11e8-b5fe-01aa75ed71a1/language-en.](#)

This study aims to realize a statistical analysis of competitiveness in the EU-28 based on a statistical reference framework, previously defined as the European Wheel of Competitiveness. This framework comprises a list of 35 indicators (EWoC indicators) related to macroeconomics, microeconomics, globalization, environment and socio-institutional aspects. The analytical methods for this study were established in line with the Statistical reference framework for competitiveness analysis in the EU-28 Member States. As a rather broad concept, competitiveness can be described by a large set of different factors and definitions. The aim of this analysis is to show possible redundancy between indicators, but also possible explanatory power. Moreover, it tests the initial assumption that the largest principal components will be sufficient to explain the variability in the observed dataset. Finally, it detects possible patterns based on available set of indicators. Results confirmed some well-known correlations. For instance, the unemployment rate is positively correlated with the indicator People at risk of poverty or social exclusion (AROPE), which is also negatively correlated with the Real GDP per capita. Nevertheless, the performed correlation analysis did not provide any substantial results regarding some other expected correlations. It has to be mentioned at this point that, the presence or the absence of correlation results between several indicators of the set of EWoC indicators, is a fact that should not be understood as a weakness of the EWoC framework but rather as a strength due to the fact that there is not redundant information. Exploring the high correlations between the aforementioned indicators and indicators by themselves, it was decided to control for the indicator 4. Concerning the results obtained with partial correlation analysis, it was evident that most of the previously captured correlation actually comes from the Gross domestic expenditure on R&D. More precisely, when controlling the correlation for the aforementioned indicator, the correlation across EWoC indicators becomes weaker or insignificant in almost all cases. For that reason, and other reasons detailed in the article, indicator 4 was not used for the latter analysis. Finally, the correlation analysis showed that indicator 34: Control of corruption could be also removed from the set of the EWoC indicators. The cross-correlation analysis explored time series correlation and showed that for the same pair of indicators, different or opposite correlation results can be obtained. This proved that countries may react differently and in several periods in time, to similar economic changes. This result provided the idea of countries clustering. In addition, obtained correlation coefficients vary from country to country. Therefore, it was of interest to look for the evidence of formation of groups of countries. In another words, the country-clusters. EU-28The performed analysis showed that the "old" EU Member States (France, Germany, United Kingdom, Italy...) form a cluster while the Eastern EU countries, i.e. "newest" Member States (Estonia, Latvia, Lithuania, Poland, Bulgaria, Romania, Croatia...) form a separate cluster. It is also noteworthy the fact that Luxembourg stands out as being very different, insofar as it forms its own cluster (Figure 1). The explored dataset consists of 29 indicators for time span 2000-2014. Therefore, it seemed logic to examine opportunities for the reduction of dimensionality or indicators' grouping. All analysis (Principal Compo-

nents, k-means cluster analysis and Gaussian mixture model and the Hierarchical agglomerative clustering analysis) showed that indicators, either as a component of one linear combination or as a member of one group, can be classified into 4 groups. It is noteworthy the fact that indicator 29: Real GDP per capita stands out as being different, insofar as it forms its own cluster.

Baldacci, E. et al. (2016). *Big data and macroeconomic nowcasting: From data access to modelling*. Tech. rep. Statistical Books. Eurostat. DOI: [10.2785/360587](https://ec.europa.eu/eurostat/documents/3888793/7753027/KS-TC-16-024-EN-N.pdf). URL: <https://ec.europa.eu/eurostat/documents/3888793/7753027/KS-TC-16-024-EN-N.pdf>.

Parallel advances in IT and in the social use of Internet-related applications, provide the general public with access to a vast amount of information. The associated Big Data are potentially very useful for a variety of applications, ranging from marketing to tapering fiscal evasion. From the point of view of official statistics, the main questions is whether and to what extent Big Data are a field worth investing to expand, check and improve the data production process and which types of partnerships will have to be formed for this purpose. Nowcasting of macroeconomic indicators represents a well-identified field where Big Data has the potential to play a decisive role in the future. In this paper we present the results and main recommendations from the Eurostat-funded project "Big Data and macroeconomic nowcasting", implemented by GOPA Consultants, which benefits from the cooperation and work of the Eurostat task force on Big Data and a few external academic experts.

Fernández-Ugalde, O. et al. (2016). *LUCAS soil component: Proposal for analysing new physical, chemical and biological soil parameter*. Tech. rep. EUR 28038EN. Joint Research Centre of the European Commission. DOI: [10.2788/884940](https://publications.jrc.ec.europa.eu/repository/bitstream/JRC102485/lb-na-28038-en-n%20.pdf). URL: <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC102485/lb-na-28038-en-n%20.pdf>.

While unemployment in the EU is above 10%, the job vacancy rate also remains high around 1.5%. This suggests considerable unmet demand for skills, which is in the focus of the EU employment promotion policies. This paper studies the special role that schooled ICT experts in firms - an intangible input often neglected and difficult to measure - play for productivity. The effects are investigated both in isolation and in conjunction with the impact of ICT maturity on microdata in six European countries (UK, France, Sweden, Norway, Denmark and Finland) for the period 2001-2009. We find that increases in the proportion of ICT-intensive human capital boosts productivity. This seems to confirm the case in favour of recruitment of highly skilled ICT employees. However, the gains vary across countries and industries, suggesting that the channels through which the effects operate are narrower for ICT-intensive human capital than for skilled human capital in general. Our findings provide an important message to the EU employment policy debate that currently revolves around the skill mismatch in general and the unmet demand for ICT skills in particular.

Boxall, M. et al. (2015). *ESS guidelines on seasonal adjustment*. DOI: [10.2785/317290](https://ec.europa.eu/eurostat/documents/3859598/6830795/KS-GQ-15-001-EN-N.pdf). URL: <https://ec.europa.eu/eurostat/documents/3859598/6830795/KS-GQ-15-001-EN-N.pdf>.

The establishment of common guidelines for seasonal adjustment (SA) within the European Statistical System (ESS) is an essential step towards a better harmonisation and comparability of infra-annual statistics, especially Principal European Economic Indicators (PEEIs). The ESS Guidelines on Seasonal Adjustment address the need for harmonisation expressed on several occasions by many users such as the European Central Bank (ECB), European Commission services, and the ECOFIN Council. The definition of best practices in the field of seasonal adjustment has been long debated at European level. Since 2007, the Seasonal Adjustment Steering Group co-chaired by Eurostat and the ECB gave a new and crucial input to the compilation of the first edition of the guidelines, published in 2009. The first edition has been widely accepted and implemented. However, taking into account the experience accumulated since 2009 and the need to further clarify some specific aspects, in 2012 the Seasonal Adjustment Steering Group decided to launch a revision of the guidelines. The ESS Guidelines on Seasonal Adjustment are the outcome of the revision work and the ESS Committee (ESSC) endorsed them in November 2014. The revised ESS Guidelines on Seasonal Adjustment present both theoretical aspects and practical implementation issues in a friendly and easy to read framework, thereby addressing both experts and non-experts in seasonal adjustment. They meet the requirement of principle 7 (Sound Methodology) of the European Statistics Code of Practice and their implementation will also be in line with principles 14 (Coherence and Comparability) and 15 (Accessibility and Clarity). The guidelines also foster the transparency of seasonal adjustment practices by encouraging the documentation of all seasonal adjustment steps and the dissemination of seasonal adjustment practices by means of the metadata template for seasonal adjustment. Finally they allow for development of expertise and capacity building. The revised version of the guidelines includes a new section with a policy for seasonal adjustment, making the revised version of the guidelines consistent with the guidelines on revisions policies. It also better describes the different steps in seasonal adjustment. Finally the specification of alternatives has been reviewed, making them more operational.

Hagsten, E. and A. Sabadash (2014). *The impact of highly-skilled ICT labour on firm performance: Empirical evidence from six European countries*. Tech. rep. JRC89703. Working Papers on Digital Economy 2014-02. Joint Research Centre of the European Commission. URL: https://ec.europa.eu/jrc/sites/jrcsh/files/ReqNo_JRC89703_The%20Impact%20of%20Highly-skilled%20ICT%20Labour%20on%20Firm%20Performance%20Empirical%20Evidence%20from%20Six%20Countries.pdf.

While unemployment in the EU is above 10%, the job vacancy rate also remains high around 1.5%. This suggests considerable unmet demand for skills, which is in the focus of the EU employment promotion policies. This paper studies the special role that schooled ICT experts in firms - an intangible input often neglected and difficult to measure - play for productivity. The effects are investigated both in isolation and in conjunction with the impact of ICT maturity on microdata in six European countries (UK, France, Sweden, Norway, Denmark and Finland) for the period 2001-2009. We find that increases in the proportion of ICT-intensive human capital boosts productivity. This seems to confirm the case in favour of recruit-

ment of highly skilled ICT employees. However, the gains vary across countries and industries, suggesting that the channels through which the effects operate are narrower for ICT-intensive human capital than for skilled human capital in general. Our findings provide an important message to the EU employment policy debate that currently revolves around the skill mismatch in general and the unmet demand for ICT skills in particular.

Pantea, S., F. Biagi, and A. Sabadash (2014). *Are ICT displacing workers? Evidence from seven European countries*. Tech. rep. JRC9112. Joint Research Centre of the European Commission. URL: https://ec.europa.eu/jrc/sites/jrcsh/files/JRC91122_ICT_displacing_workers.pdf.

This paper examines whether ICT substitute labour and reduce the demand for labour. We used firm-level comparable data separately for firms in manufacturing, services and ICT-producing sectors from seven European countries. We adopted a common methodology and applied it to a unique dataset provided by the ESS-Lait Project on Linking Microdata. We controlled for unobservable time-invariant firm-specific effects and we found no evidence of a negative relationship between intensity of ICT use and employment growth. We read this as an indication that ICT use is not reducing employment among ICT using firms.

Sabadash, A. (2014). *Employment of ICT specialists in the EU (2000-2012)*. Tech. rep. JRC92503. Working Papers on Digital Economy 2014-01, MPRA Paper 61644. Joint Research Centre of the European Commission. URL: https://ec.europa.eu/jrc/sites/jrcsh/files/JRC92503_Employment_of_ICT_Specialists.pdf.

This study examines the evolution of the number of ICT-skilled workers employed in industry sectors in the EU28 over the period 2000-2012. Data are taken from the Eurostat Labour Force Statistics. It introduces a novel definition of ICT specialists that combines occupations and skills taxonomies. For the period prior to the introduction of the Standard Classification of Occupations (ISCO-08) it starts from the OECD definition but includes a wider range of ICT occupations. From 2011 onwards it adopts the thematic view for ICT occupations proposed by the ILO (2012). It confirms that employment of ICT specialists in the EU27 has been resilient to the economic downturn and uncertainty in global labour markets, and was able to maintain a growth path of 4.3% per year over the period 2000-2012, more than 7 times higher than average growth of total employment over the same period. Though ICT employment evolved cyclically it never turned negative. This rapid growth in ICT employment confirms the increasing importance of ICT technologies in the global economy.