EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS

Q 2022 VILNIUS

ABSTRACTS

Vilnius, Lithuania
8–10 June 2022

Welcome to Q2022 Conference

Dear colleagues,

Statistics Lithuania and Eurostat are pleased to welcome you to the European Conference on Quality in Official Statistics Q2022, held in Vilnius on 8–10 June 2022. The event is preceded by a series of training courses on 7 June (one-day training courses on quality management, the role of statistics in the era of Big Data and in a future society, innovation and modernisation practices, as well as data journalism and data visualisation).

The Q2022 conference is the tenth consecutive quality conference involving the international statistical community.

The conference aims to bring together statisticians, academics and stakeholders to reinforce the dissemination of knowledge and recent research on emerging questions related to quality in official statistics. Furthermore, the conference aspires to foster cooperation between users and producers, to enhance cooperation and exchange between official statistics and the scientific community, and to provide a forum for the exchange of good practice in meeting new challenges in official statistics.

The event will also focus on the institutional challenges of quality management, quality assurance in the emerging data ecosystem, and new challenges in creating and maintaining trust among users, as well as new opportunities in the face of global crises. This conference should also draw the attention of governmental bodies to the importance of high-quality, timely and more detailed statistics, and increase awareness among other stakeholders of the challenges faced by producers of official statistics, especially in times of crisis.

This event will benefit from sharing your ideas and discussions on topical issues for the benefit of all the statistical family.

Dr. Jūratė Petrauskienė
Director General of Statistics Lithuania

Dr. Mariana Kotzeva
Director-General of Eurostat

Statistics Lithuania, **Statistics Lithuania**

Statistics Lithuania is a public authority that participates in developing and implementing public policy in the field of statistics assigned to the Minister of Finance, and coordinating official statistics in the country pursuant to the provisions of Part I of the Official Statistics Programme.

The Mission of Statistics Lithuania is to prepare and disseminate official statistics necessary for decision-making and to inform the public about issues relevant to the country as well as to create and develop the State Data Governance Information System.

Our strategic objective is timely production and provision of objective and reliable statistical information with regard to changing needs, and consolidation of the data required for the State's information needs in the State Data Governance System by ensuring their quality and security.

Eurostat, **eurostat**

Eurostat produces European statistics in partnership with National Statistical Institutes and other national authorities in the EU Member States. This partnership is known as the European Statistical System (ESS). It also includes the statistical authorities of the European Economic Area (EEA) countries and Switzerland.

Eurostat coordinates statistical activities at Union level and more particularly inside the Commission.

Scientific Committee

Dr. Mariana Kotzeva, Co-Chair, Director General, Eurostat
Dr. Jūratė Petrauskienė, Co-Chair, Director General, Statistics Lithuania
Dr. Dominik Rozkrut, President, Statistics Poland
Jean-Pierre Poncelet, Director, Eurostat
Dr. Gabriela Vukovich, President, Hungarian Central Statistical Office
Dr. Aurel Schubert, Chair, ESGAB
Dr. Biruta Sloka, ESAC
Dr. Silke Stapel-Weber, Director General for Statistics, ECB
Steven Vale, Regional Adviser in Statistics, UNECE
Dr. Misha Belkindas, President, IAOS
Dr. Kazimieras Padvelskis, Vilnius Gediminas Technical University, Lithuania

Programme Committee

Claudia Junker, Co-Chair, Eurostat
Dr. Inga Masiulaitytė-Šukevič, Co-Chair, Statistics Lithuania
Dr. Jean-Marc Museaux, Eurostat
Véronique van der Zande, Eurostat
Dr. Vilma Nekrašaitė-Liegė, Statistics Lithuania
Thomas Burg, Statistics Austria
Monika Bieniek, Statistics Poland
Kornélia Mag, Hungarian Central Statistical Office
Giorgia Simeoni, ISTAT
Maria João, Statistics Portugal
Dr. Audronė Jakaitienė, ESAC
Dr. Gindra Kasnauskienė, Vilnius University, Lithuania
Dr. Domantas Jasilionis, Max Planck Institute for Demographic Research
Remi Prual, Estonian Quality Association

Key speakers

Dr. Dominik Rozkrut, President, Statistics Poland
Anil Arora, Chief statistician of Canada

Content

Special session 1
ESSNET SMART SURVEYS: CURRENT CHANCES AND CHALLENGES

Title of abstract: Trusted Smart Surveys: Solutions for the European Statistical System – Objectives and Main Challenges

Shari Stehrenberg – Destatis, Wiesbaden, Germany
Konstantinos Giannakouris – Eurostat, Luxembourg, Luxembourg

(Trusted) Smart Surveys offer opportunities to improve the quality of survey data and reduce the response burden by combining traditional survey data collection methods with non-traditional and new forms of data collection. However, Smart Surveys may run the risk of not being very successful in terms of adoption and acceptance by the respondents that are data providers and users of smart devices, due to concerns about privacy. Mitigating this risk, a carefully designed and cleverly communicated Trusted Smart Surveys system has better chances to be adopted and accepted. This should be considered on top of the maturity of the smart technologies, the development of robust methodologies and the necessary data quality, which are all critical elements to user engagement and adoption of the smart device and sensor paradigm and services. The ESSnet Smart Surveys project aims to investigate the development and dissemination of (Trusted) Smart Surveys by designing a European platform to share Trusted Smart Survey solutions and components.

Smart Surveys encompass (inter)active data explicitly provided by respondents as well as passive data collected in the background by sensors in smart devices (e.g. smartphones, activity trackers). Trusted Smart Surveys refer to an augmentation of the Smart Survey concept by technological solutions aimed at increasing the degree of trustworthiness, hence promoting public acceptance and participation. Constituent elements of Trusted Smart Surveys are the strong protection of personal data based on privacy-preserving computation solutions, full transparency and auditability of processing algorithms.

The project encompasses four pilot studies and experimental variations. The two larger pilot studies are related to the Time Use Survey and the Household Budget Survey. Both surveys have been chosen as they are considered to be burdensome to respondents and hence to be prone to low recall as well as underreporting errors. Modernisation of these surveys – as well as similar ones – is therefore urgently needed. In addition, two smaller pilot studies are carried out in the areas of health and living conditions, using activity trackers and indoor air-quality measurement sensors.

To employ the potential of (Trusted) Smart Surveys, both methodology and architecture of statistical processes need to be modified and extended. Therefore, the project aims at conceptualizing a European platform supporting the use of shared smart survey solutions and to assess the usage of applications for European social surveys. For this, a conceptual framework including methodological, technical logistical and legal dimensions of such surveys will be investigated.

In this presentation, we will provide an overview of the objectives of the ESSnet Trusted Smart Surveys project and introduce the work packages. We will address the main obstacles of Trusted Smart Surveys as solutions for the ESS and give an outlook on future work related to Trusted Smart Surveys.

Keywords: Trusted Smart Surveys; privacy enhancing technologies; digital devices; data collection, Trusted Smart Statistics.

Title of abstract: The Trade-off between Active and Passive Data Collection in Smart Surveys. Four Case Studies

Annemieke Luiten, Barry Schouten – CBS, The Netherlands
Patrick Lusyine – StatBEL, Belgium

Smart surveys face an important question in designing and analyzing the combination of survey questions, sensor measurements and data donation: To what extent should respondents actively be involved in data collection, in particular sensor measurements? The rationale to employ sensor data and other forms of data in surveys is to reduce burden for the respondent, and/or to improve measurement quality of data. However, these data are subject to inaccuracy as well. They may have gaps and they may have both systematic and random measurement error. Furthermore, the data quality may itself depend on the motivation and skills of respondents, if they are to initiate the measurements. There are, therefore, three motives for involving respondents: Improving data quality by letting respondents check or validate data, increasing involvement and motivation by making them engaged in various steps, and increasing transparency with for example GDPR in mind. This leads to a complicated trade-off between data quality and respondent burden.

In the paper, the ESSnet Smart survey case studies on consumption, time use, health and living conditions are discussed. In all, the extent of active respondent involvement plays a prominent role: receipt scanning in household budget surveys, location tracking in time use, physical activity tracking in health and indoor climate measurements in living conditions.

Keywords: smart surveys; smart devices; data collection; machine learning; response rates.

Title of abstract: Smart survey trade-offs between legal requirements, quality control and in-device processing

Barry Schouten – Destatis, Germany
Sulaika Duijsings-Mahangi – Ministry of Infrastructure and Water State, The Netherlands

Smart surveys collect and combine new forms of data. These forms of data offer opportunities to reduce respondent burden and measurement error, but also contain new sources of error. Handling some of the errors requires interaction with the respondent and/or adding other sources of data in-house. Doing so a potential friction arises with GDPR/privacy legislation, because both detailed data and quality metadata may have to be sent to the statistical institute or data has to be sent to the respondent device. While GDPR is relatively clear on handling 'perfect' data, it is less clear on handling 'imperfect' data and quality metadata. ESSnet Smart Surveys invited a working group of NSI legal officers and survey designers. Over the course of a year, this working group considered a number of scenarios as a function of potential smart features in a survey and identified a number of main open questions.

In the paper, we discuss the working group considerations, scenarios and potential solutions based on the ESSnet case studies.

Keywords: smart survey; quality control; GDPR.

Title of abstract: The Methodological and Architectural Framework for the European Platform for Trusted Smart Surveys

Mauro Bruno, Claudia De Vitiis – ISTAT, Italy
Jacek Maslankowski – Statistics Poland, Poland
Niels Meise – Destatis, Germany
Joeri van Etten – CBS, The Netherlands

Trusted Smart Surveys are surveys in which respondents employ smart devices to collect survey data through active and passive data collection and can share existing data collected by trusted third parties, like government authorities and larger, stable enterprises willing to establish data delivery agreements. This innovative type of data collection offers new challenges to improve the quality of social surveys for the NSIs. The ESSNet on Smart Surveys, which started its activities at the beginning of 2020, is delivering preparatory work to create a European wide platform to share and re-use smart survey solutions and components.

In this context, the work-package 3 is working on the idea of TSS platform through the conceptualization and development of a methodological and architectural framework for trusted smart surveys, following a top-down design approach. The platform will be agnostic to particular application domains, flexible and modular, implementing a set of common functions and configurable services that can be used to build particular instances of trusted smart surveys for specific application domains and/or target areas.

The target architecture will model smart statistical processes resulting from the combination of traditional and new data sources (e.g. sensor data, geolocation). As part of the definition of the TSS platform, a specific line of work is dedicated to the specification of the metadata that are useful in the context of smart surveys and special attention will be given to data structures and flows, process execution and traceability, and privacy, together with the representation of sensor data.

The activity of the work-package is also focused on the development of Proof-of-Concepts, working in particular on machine learning, technical infrastructure, privacy preservation, metadata and architecture, through the identification and implementation of use cases.

Keywords: smart surveys; smart devices; architectural framework; machine learning, sensor data.

Special session 2
IMPLEMENTATION OF QUALITY MANAGEMENT FRAMEWORKS SYSTEMS IN THE EUROPEAN NEIGHBORHOOD POLICY (ENP) SOUTH COUNTRIES

Title of abstract: The Methodological Changes in Data Collection Modes and Alignment of Quality Tools for Statistical Operations implemented in COVID-19 Pandemic Period – PCBS Experience

Lena Qadi, Nayef Abed – Palestinian Central Bureau of Statistics (PCBS), Ramallah-Palestine

The Covid-19 pandemic has casted its shadows on Palestine at the since March 2020, and it began to spread widely, so a complete lockdown was imposed for more than a period in addition to imposing severe Precautions and preventive measures. During this period, there was a disruption of fieldwork based on face-to-face data collection. While Palestinian Central Bureau of Statistics (PCBS) had to provide the necessary data to meet the needs of policy and decision-makers from the public and private sectors, civil society and international institutions in a way that contributes to developing programs and interventions that can mitigate the effects and consequences of this pandemic.

PCBS worked on developing new tools and methodologies that comply with the stage requirements, the most important of which was adapting totally data collection by the phone (CATI) as an alternative method of face-to-face data collection interview (CAPI), reconsidering the plan of the statistical program for the year 2020, setting priorities for statistical activities, focus more on improving the administrative records data from official organizations to be the main source for some statistical indicators wherever possible, and benefiting from all available periodic data to do forecasting and projections.

In view of the mentioned methodological changes, it was necessary to align the quality tools in order to be valid for measuring quality dimensions indicators according to the new approach in data collection modes, especially that all previous quality procedures and tools were prepared according to the in-person interview modes.

This paper will present PCBS experience in the procedures of data collection during the Covid-19 Pandemic from many aspects. In particular, the impact of changing the data collection mode to phone (CATI) and the challenges it faced, especially the lack of Sampling Frames that include accurate phone numbers of households. We needed to do many comparisons to be sure that the responded samples were not biased or under-represented. In addition, the results of some quality indicators for surveys carried out during that period will be added.

Keywords: Data Collection Modes, Covid-19 pandemic, Quality tools, quality indicators, PCBS.

Title of abstract: The implementation of a quality management system strategy: Digital transformation as an accelerating factor

Sekkak Safaa – HCP, Rabat, Morocco

Nowadays, there is a widespread interest in Quality for the High Commission for Planning of Morocco (HCP)as a whole which is the main reason of the implementation of a quality management system. The paper will outline the organizational aspects to engage and motivate HCP staff on Quality subjects as well as the set up of a strategy of how to ensure that the data production system in HCP is following the Quality Management requirements as defined in many standards. We will also highlight the relationship between the digital transformation implementation plan and the quality management as a key factor helping the quality to be an essential component of HCP performance. Major results include: the adoption of the quality management approach is highly correlated with the management of the digital transformation.

We are focusing as a starting point on the strategy of the organisation and the implementation of a quality management system in HCP by using Generic Statistical Business Process Model (GSBPM) as a representation to define and describe statistical processes in a coherent and standardised way, not only to identify the gaps in the existing processes that needed to be filled but also to serve as the basis to evaluate the results and the progress of our action plan and to establish the Quality Audits and evaluation using GSBPM.

Regarding engagements with the data users, HCP is adopting SIMS (single integrated metadata structure) for its surveys as formal strategic documents and plan to cooperate with other producers of statistics to ensure the best practices in reporting on quality and by this spreading quality management culture among line ministries and other entities in the national statistical system in Morocco.

Keywords: Quality management system; Digital transformation; GSBPM; data ecosystem.

Title of abstract: The role of official statistics as a reliable source of information in the data society

DR. Shimaa Mostafa Gaber Ibrahim – Central Agency for Public Mobilization and Statistics (CAPMAS), Cairo, Egypt

Official statistical data plays an active role in all economic, political and social fields. So, coordination is made between all statistical agencies worldwide to apply all standards in conducting and implementing surveys and censuses to ensure their effectiveness in the data community. The study aims to insure that applying all International Monetary Fund standard lead to improve the production of official statistical data.

The study is concerned with distinguishing between the General Data Dissemination System (GDDS) and the Special Data Dissemination Standards (SDDS). The study questions revolve around the extent of unification of concepts and applications used in various statistical topics.

Keywords: statistical agencies, Data Dissemination System.

Title of abstract: New trends and challenges of creating and maintaining trust of users in Jordan

Duraid Alshawawreh – Department of Statistics, Amman, Jordan

The Department of General Statistics operates under the Jordanian Statistics Law No. (12) of 2012, and is considered the main driver of the Jordanian statistical system. It always seeks to establish standards for quality assurance and control in the department's operations and components of the statistical system, and to enable the Department to develop produce data, and to improve methodologies and means of data production, dissemination and increasing statistical awareness in effective partnerships with all national institutions; by keeping abreast of new trends and challenges related to establishing users trust.

The most important categories of data users are policy-makers, decision-makers, investors, scholars and researchers. This requires knowing the type of data required and determining it, its priority and its sources. Therefore, the department maintains relations with them to exchange needs and opinions in order to obtain their feedback aimed at continuous improvement. And improving the level of data users' satisfaction by preparing a policy for disseminating and raising awareness of statistical data, designing an annual, semi-annual and quarterly community opinion poll, and preparing an improvement plan from the community opinion report. This is according to the community's need for economic, social, demographic indicators and any indicators of importance, and specifying data according to its objective and priority based on plans and projects at the national level, consultations with data users are conducted based on the need and the proposed projects.

Hence, the department has formed technical committees of specialists in the department and the main partners from the partner institutions, assessing the achievement of the statistical situation and seeking the assistance of international experts from abroad in order to improve its methodologies, operations and activities according to the results of following up on data users satisfaction, and focusing on managing data and quality in an efficient manner and maintaining its sustainability, and that By storing it in correct ways and determining the powers of access to that data within the internationally applied quality control standards.

Keywords: Partnerships, Users trust, Data management, Statistical communication, Quality assurance.

Title of abstract: Automation of the quality control process for data collection

Moncef Jarboui – National Institute of Statistics, Tunisia

INS Tunisia was integrated the use of new technology for data collection since 2015, this integration involves improving the quality of the statistical process specially securing data collection via mobile devices.

To achieve this goal, INS Tunisia designed a modular system based on private cloud, mobility management, real-time control of collected data and automatic coding. INS data center hosts the private cloud enabling access to servers, data and applications anytime, anywhere, and guaranteeing granular access control to protect sensitive data while facilitating productivity.

To ensure the security of the mobile fleet and cover the management of mobile devices, applications and data, INS uses the Enterprise Mobility Management (EMM). This EMM provides a virtual private network (VPN) for each application encrypting data on an application-by-application basis rather than encrypting all data exchanged between field devices and internal INS servers. Moreover, it allows automatic and remote installation or updating applications without downtime by user profiles linked to separate applications on their devices.

To supervise and control data collection in real-time, web applications are developed for each survey and used by different profiles (master user, regional office manager and supervisor).

Considering the importance of coding in the data collection process, a new coding system is established to provide automatic coding of business activity and profession in demographic surveys. This system is based on three main objectives:

– Prepare the knowledge bases for the activity and profession variables
– Develop different applications: the first to prepare the knowledge bases, the second to codify automatically these variables and the third to codify the non-coded records (the residue)
– create experts' team in the automatic coding.

Keywords: data collection via mobile devices; coding; demographic surveys.

Special session 3
IMPLEMENTATION OF QUALITY MANAGEMENT FRAMEWORKS SYSTEMS IN THE ENLARGEMENT AND EUROPEAN NEIGHBORHOOD POLICY (ENP)-EAST COUNTRIES

Title of abstract: Quality Management System for official statistics of Montenegro

Bojana Radojević, Dragana Zivkovic – MSc, Statistical office of Montenegro, Podgorica, Montenegro

This article describes the long-term process of development and implementation of the Quality Management System in the official statistical system of Montenegro. Work on the implementation of quality begins in 2010 with the conducting of the Adapted Global Assessment of the National Statistical System of Montenegro to the establishment of the Quality Management System. The AGA report provides recommendations for improving the new law on statistics. Most of these recommendations from the report are included in the new Law on Official Statistics and the Official Statistics System.

At the initiative of the MONSTAT, the Light Peer Review – LPR was conducted in 2011, after which a comprehensive quality management plan was defined, with the strong support of the newly adopted law on statistics. The establishment of the quality infrastructure was based on the core elements of a Quality Management System (Quality Policy and Objectives, Quality Manual, Organizational Structure and Responsibilities, Data Management, Processes, User Satisfaction, Continuous Improvement, Quality Instruments and Document Control and Management). The Peer Review in 2018 confirmed that MONSTAT enjoys a high level of trust and its independence is largely recognized.

Today, a decade later, MONSTAT is able to monitor the development and changes in the quality of statistical production with the help of quality indicators, achieve continuous quality improvements in statistical surveys and standardize statistical production processes by applying the GSBPM quality model. It presents to users the production of statistical data through the available METADATA system and monitor the progress of all producers of official statistics through quality reports.

By establishing self-assessment, MONSTAT monitor the development of statistical production through three components (institutional environment, statistical results and statistical processes). MONSTAT use the documentation collected during the assessment to identify weaknesses and identify areas for future improvement.

Our future step is (1) stronger support of other producers of official statistics that produce European statistics in the implementation of quality management in order to improve the overall quality of official statistics in Montenegro and (2) continuous quality improvement.

Keywords: Quality Management System; Peer Review; Self-assessment; GSBPM.

Title of abstract: Institutional challenges of quality management in the State Statistical Office of the Republic of North Macedonia

Mira Todorova – State Statistical Office, Skopje, the Republic of North Macedonia

The paper aims to present the latest activities in State Statistical Office that contributed to strengthen institutional capacity for quality management.

The notion of "quality" is undoubtedly complex and multidimensional. The production of high-quality statistics is part of SSO mission. In the past, the institution has undertaken different activities to improve quality of products and processes, but these were separate activities.

Under IPA 2015 and 2017 Programme, SSO participated in the project for Quality management. The main objective of these projects was to assess present quality management system in NSIs and to assist establishment of quality management system and improve quality reporting. A working group has been set up in order to implement IPA projects.

The actions taken in both projects have complemented each other and covered quality management and quality measurement. In the paper will be presented the results of the projects and how they strengthen quality management in the institution.

As a result, SSO has defined the milestones of quality management, produced several strategic documents related to quality and introduced as part of regular process production of SIMS reports. The internal quality audit was set up and one survey was audited.

Last but not least, is the communication of quality among staff and promotion of importance of quality for branding "official statistics".

Our work on quality was based upon two milestones: EC Code of Practice and recommendations from last Peer Review for SSO.

Further improvement of quality management is expected under IPA 2019 programme for Quality management.

Keywords: Quality management; quality framework; SIMS reporting; internal quality audit.

Title of abstract: On the Road of QMS – Bumps and Speedups

Nataša Cvetković, Miodrag Cerovina – Statistical Office of the Republic of Serbia, Belgrade, Serbia

Statistical Office of the Republic of Serbia (SORS) has been implementing a systematic approach to quality assessment, quality assurance and quality reporting by using different methods and tools under the umbrella of the ESS recommendations. One of the important strategic goals of the SORS in past and recent years was to establish a comprehensive quality management system (QMS) to be able to continuously monitor, evaluate and improve the quality of its statistical products and processes as well as a level of overall performance. A small but agile quality team is committed to the systematic work on introducing the quality management framework, methods and tools with a policy of continuous improvement and permanent reconsideration on what had been done. Through describing obstacles, but also ways to overcome them, this paper will show the specifics of the Serbian vision of its comprehensive QMS, already implemented elements, as well as the work ahead.

Keywords: quality management system; quality framework; quality methods and tools.

Title of abstract: Implementation of quality indicators in the national model of activities of state statistics bodies

Vadym Pishcheiko – State Statistics Service of Ukraine, Kyiv, Ukraine

The State Statistics Service of Ukraine is working under the creation a reliable and effective quality management system.

In 2010, the Principles of State Statistics were developed, which are harmonized with the European Statistics Code of Practice, after amendments to which in 2017, the SSSU in 2018 developed a new version of the Principles of State Statistics.

The State Statistics Service developed and approved in 2016 the Policy on Quality in State Statistics Bodies, in 2019 the Risk Management Policy of State Statistics Bodies.

In 2018, in accordance with the standard model of statistical business processes (GSBPM), the National Model of Statistical Production and technological maps of processes and sub processes of statistical production were developed.

Since 2019, the SSSU has introduced risk management, since then annually identifies and assesses risks, develops measures to eliminate them and reduce the negative impact on the activities of state statistics, monitors the implementation of these measures.

In 2020, on the basis of the standard model of activity of the statistical organization (GAMSO), the Description of the national model of activity of state statistics bodies was prepared, which includes quality indicators. The preparation of the Inventory

provided an opportunity to improve the classifier of works, which is used in the technological program of state statistical observations and in the automated system of accounting for working time.

In the second quarter of 2021, a survey of the quality management system of the State Statistics Service was conducted in accordance with the Quality Assessment Framework (QAF). A similar survey was conducted in 2019.

In 2021, a test monitoring of the Questionnaire was conducted in order to check quality indicators according to the processes of the Description of the national model of activity of state statistics bodies based on the results of activities in 2020.

The Commission for coordination of the quality management system of statistical activities of the State Statistics Committee was established and the Plan-schedule of the Commission's work in 2021 was developed and approved.

In 2021, the Commission considered and approved:

– regulations on the head of the process of the quality management system of the State Statistics Service;
– the structure of the description of the documented process of the quality management system of the statistical activity of the State Statistics Committee;
– system of monitoring the quality of the processes of activity of state statistics bodies.

Keywords: quality; system; process; indicators.

Title of abstract: Key elements of the quality management system at Armstat

Anahit Safyan – Statistical Committee of the Republic of Armenia, Yerevan, Republic of Armenia

The Statistical Committee of the Republic of Armenia (Armstat) will present a paper on its experience on how the quality work has been organized and the quality management system has been established, and what are the key elements of this system. The quality management system is a necessary tool for development, production and dissemination of official statistics and for maintenance of its exclusive position and mission to provide public with high quality and evidence-based statistics. This exclusiveness is anchored on the UN Fundamental Principles of Official Statistics and European Statistics Code of Practice, such as professional independence, scientific methods and a commitment to quality, the implementation of which gives exceptional value to official statistics and increases users trust.

Important prerequisites for quality assurance are the legislative mandate of producers of official statistics and the obligation to use international principles and standards.

Paper will discuss the quality system and standards used. Key elements of the quality management system are institutional infrastructure, quality policy document, European Statistics Code of Practice, Quality Assurance Framework of the European Statistical System and Peer Reviews. Standards used are quality declarations developed for 138 statistical products, the GSBPM (Generic Business Process Model)-based documented statistical business processes for 138 products (visualized statistical processes using the so-called "swimming pool model"), and the self-evaluation questionnaire for complying with the European Statistics Code of Practice for internal quality audit. Paper will cover the work on harmonizing quality reporting and metadata by introducing reference metadata reporting standards: EU SIMS (Single Integrated Metadata Structure) and reporting formats ESMS (Euro SDMX Metadata Structure) and ESQRS (ESS Standard Quality Report Structure).

The reference metadata system supports quality reporting and is also used for metadata dissemination in a centralized and standardized manner to provide high quality, timely and internationally comparable statistics.

Keywords: Quality management, quality framework, quality declaration, GSBPM, SIMS.

Special session 4
FROM EXPERIMENTATION TO PRODUCTION

Title of abstract: Machine learning for coding occupations in the Census: first lessons from experiment to production

Elise Coudin, Aurélien Fortin, Théo Leroy, Lucas Malherbe, Tom Seimandi – INSEE Montrouge, France

Occupational classifications are useful tools used by statisticians, economists, sociologists to provide descriptors both accounting for similarities in job tasks and contents and similarities in economic and institutional contexts. To provide realistic social or economic analyses, occupational classification dictionaries have to be regularly updated. In 2020, a new dictionary of the French occupation classification (PCS 2020) was disseminated, accompanied with an autocompletion tool, which links perfectly a list of 5,000 jobs to their classification category. Only responses not in this list remain to be coded. INSEE has chosen not to adapt its rule-based automatic coding system set to code within the previous dictionary (PCS 2003) to the new dictionary. INSEE rather has chosen to experiment the use of machine learning techniques to perform this type of classification task for which they are expected to perform well. In 2021, a large campaign of manual labelling was conducted: around 100,000 Census job answers were labelled in PCS 2020, each twice, by two different manual coders, and a third arbitrage when required, with the aim of ensuring the quality of the training/test sets on which the algorithms would be trained/tested. A two-layer neural network algorithm (FastText embeddings of n-grams and classifier) was finally selected. The experiment suggests that the combination of the two automatic coding modes (list and supervised learning on non-lists) allows to reach or even exceed the accuracy rates of the previous system at the finest level for the current occupation, but not for the previous occupation (retired and unemployed) which has more paper slips. The combination with a part sent to manual work allows to gain some points of accuracy.

Based on these results, the integration of predicting and training tools into the Census production chain is investigated during the first quarter of 2022, with the aim of having the 2024 Census campaign coded in PCS 2020. This covers evaluating costs and gains of the integration of (part of) the modules developed during the experiment. This covers defining the new organization of the Census production relative to occupational coding, defining different roles and strategy to evaluate and control the quality of coding by the algorithms. This covers also keeping as optimal target another, much more ambitious, challenge which is the construction of a completely mutualized tool to code in PCS 2020 data from different sources and for different actors.

Keywords: Experimentation, occupational classification, NLP, supervised machine learning, automatic coding, fastText.

Title of abstract: All statistics great and small – experiences with innovation in Dutch statistics

Barteld Braaksma – CBS, The Hague, The Netherlands

Innovation is an ever-ongoing process in society and not less so in official statistics. It is quite clear that the way statistics are produced is evolving continuously, depending on new sources that become available, new methods that are being developed and new requirements articulated by users. Examples of new sources are aerial pictures, scanner data, signalling data from mobile phones, social media, VAT (value added tax) declarations and websites. New methods are e.g. machine learning, primary data collection using mobile devices, privacy-preserving techniques, small area estimation and web scraping. New user requirements fall apart in two broad categories. On the one hand, users call for new statistical information on emerging societal phenomena like digitalisation, globalisation and sustainability. On the other hand, users ask for better accessibility of existing statistics through e.g. dashboards and infographics, machine-readable open data and access to synthetic datasets. Most developments that may trigger innovation in official statistics come from outside the statistics community and cannot be handled by statisticians alone. Increasingly, statistical institutes have to engage in collaborations with partners from academia, government and the private sector- both to improve and to accelerate innovation. Developing successful collaborations may in itself depend on innovative ways of co-creation.

By its very nature, innovation is not always easy. It may affect business continuity and have an impact on quality. New skills and additional resources may be required. Challenges and risks for incremental innovations can be limited, but for more disruptive or radical innovations they may be quite significant. Over the years, CBS has gained ample experience with innovation in different forms, both in large programs and in small projects. An example of the former is the HEcS+ program, that took on a full redesign of the chain of economic statistics. HEcS+ ran for several years and simultaneously led to burden reduction, cost saving and quality improvement. An example of the latter is a project to measure the impact of the internet on Dutch economy. In itself this was a one-off experimental statistics project but it led to different spin-offs, including a project to link website information routinely to the CBS statistical business register. Of course, not all innovative activities have been successful. For example, the use of Dutch mobile phone data to support the fight against the COVID pandemic with relevant statistics has so far not proven possible, even though all technical and methodological preparations had been done.

In this paper we discuss enablers and obstacles for innovation, illustrated by concrete examples.

Keywords: experimental statistics, innovation, open data, big data, synthetic data.

Title of abstract: Building Big Data-based production pipelines: reflection on quality requirements and issues

Fabio Bacchini, Monica Scannapieco – Istat, Rome, Italy

Experimental Statistics have been promoted by Eurostat as statistics that "have not reached full maturity in terms of harmonisation, coverage or methodology, they are always marked with a clearly visible logo and accompanied by detailed methodological notes" [1]. Experimental statistics are characterized by new sources, such as Big data, able to improve several quality aspects including timeliness. However, the use of Big data claims for a new challenge in the quality documentation for which a well-defined framework is not implemented yet.

Istat, the Italian National Statistics Institute, publishes experimental statistics, based on the use of new sources as well as on innovative production methods, in a dedicated part of its institutional website [2]. Among the first experimental statistics published by Istat, the Social Mood on Economy Index (SMEI) is a daily index computed starting from the Italian Twitter's public stream aimed at representing the perception on the evolution on the economics features. More specifically, we use a keyword-based filter to select those tweets that are considered as pertinent to the economy and based on a lexicon, we assign pre-computed sentiment scores to words composing the selected tweets. Then, a clustering step permits to classify each single tweet as positive, negative or neutral and the SME index is computed as a measure of the distribution of the scores of the tweets belonging to the positive and negative classes. Though computed on a daily basis, the index is published since October 2018 each quarter, with the daily time series provided as an attachment to the publication.

The methodological characteristics of SMEI make it a perfect candidate to investigate the quality issues related to the use of Big data. First of all, its intrinsic multivariate approach implies for difficulties in the interpretation that it is expected to evolve according to small changes in the filter fine-tuning. How are we able to track its quality characteristics? What are the specific uses of the SME index and what stakeholders is necessary to engage in order to build a production pipeline that could meet (new) users' need and benefit from shared methodological thoughts?

The paper will report our current discussion and preliminary solutions that we are going to implement. We hope it will contribute to the ESS debate on setting up quality standards for processing Big data-based statistical products.

Keywords: big data; social media; experimental statistics.

Title of abstract: From cradle to production – measuring the collaborative economy

Simon Bley, Christophe Demunter – Eurostat, European Commission, Luxembourg City, Luxembourg

The paper will present experiences made by the authors in the establishment of new experimental statistics on short-term holiday rentals offered on international online platforms. This project is one of the first practical attempts in official statistics at EU level to augment existing traditional data sources by accessing privately held data instead of relying on surveys or registers. As opposed to these classical sources of official statistics, the respondents are not the statistical units themselves, but intermediate data holders that collect the data for non-statistical purposes.

Due to this fact, establishing relevant and reliable statistics based on privately held data poses unique challenges to us statisticians. The paper outlines some of the concrete quality challenges that Eurostat needed to overcome in order to arrive at the first data release in June 2021, as well as issues that will remain to be solved in order to integrate the statistics into the realm of European Statistics. The two main open issues that will be discussed are a potential double counting of bookings between existing traditional tourism surveys and the new project, as well as a double counting of listings between the various platforms. The former prevents Eurostat from offering indicators c ombining new data with existing tourism statistics, while the latter needs to be solved in order to provide accurate capacity statistics to address current policy issues.

Keywords: privately held data; B2G data sharing; tourism; webscraping.

Special session 5
THE NEXT ROUND OF PEER REVIEWS IN THE EUROPEAN STATISTICAL SYSTEM – HOW CAN THEY HELP TO ASSURE THE QUALITY OF AND INCREASE TRUST IN EUROPEAN STATISTICS

Title of abstract: The thirds round of ESS peer reviews – a strategic view on the design and purpose

Claudia Junker – Eurostat, Luxembourg

The third round of peer reviews in the ESS has started in June 2021 and will be completed in June 2023. It is based on the revised European statistics Code of Practice. The methodology for implementing this third round of peer reviews has been developed in very close cooperation with the ESS members, in the spirit of a true peer approach. The approach taken by the ESS was to start with a strategic, high-level discussion on the objectives, scope and purpose of the peer reviews and a discussion on what the ESS would like to achieve with this new round of peer reviews. This discussion was very necessary because the objectives and the purpose of the peer reviews have an impact on other elements of a peer review such as the scope, the composition and qualifications of the peer reviewer teams' and the communication elements of a peer review.

The paper will describe the strategic reflections behind the design of the third round of peer reviews as well as some key elements of the methodology and their links to the implementation arrangements, such as e.g. the impact of improvement recommendations going beyond the authority of the NSI on the way the ESS will communicate on the peer reviews, the way the combination of a peer review and audit-like approach impacts on the formulation of recommendations and others. Communication on the peer reviews will also be used to promote the value of and trust in European statistics and the paper will elaborate on the means and tools of such a communication campaign.

Finally, the paper will also describe the reflections behind the development of the tools and instruments for implementing the peer reviews, the collaborative approach taken to develop them, the decisions on the various methodology elements and links between them.

Keywords: peer reviews, quality, communication.

Title of abstract: The impact and inspirations of the Peer review on the new mandate of Statistics Lithuania

Gita Literskė, Inga Daunaravičienė – Statistics Lithuania, Vilnius, Lithuania

The European Statistics Code of practice (ES CoP) is the backbone of the ESS common quality framework and National statistical institutes (NSIs) have committed themselves to adhere to the CoP. The compliance with the Code of Practice is periodically assessed by means of peer review mechanism.

The third round of European Statistical System (ESS) Peer Reviews aimed at improving quality and trust in official statistics by assessing the compliance with the ES CoP and based on indicated future-oriented recommendations to help NSIs to improve National statistical systems (NSS) and subsequently revise their mandates amending their legal framework. Lithuania was among the first Member States reviewed and the first with a peer review online visit.

By continuously modernising and innovating the NSS we not only improve the quality of our products and services, but also strive to more efficiently meet user needs. Envisaging this, Statistics Lithuania (SL) in 2020 legally became the owner of the State Data Governance System, which empowers wider purposes of use than official statistics. SL became like an actor of two roles: producer of official statistics, and at the same time SL empowered by the right to collect data and produce information other than official statistics.

Still essential contradictions and obstacles in terms of data re-use remain. Obviously, becoming a state data steward is not purely a technological problem, but primarily a legal issue, linked with collaboration efforts. In 2021, SL initiated fundamental changes to the national legal framework.

The article will discuss how the Peer Review recommendations can serve strengthening the NSS through processes, collaborating with stakeholders and partners and thus, amending the national legal framework. The topical issue is how to manage the new dual mandate – remaining a trusted producer of official statistics and also becoming a trustful provider of data services to support evidence-based decisions and serve as the public good.

Keywords: European Statistics Code of Practice, National Statistical System, Peer Review, legal framework.

Title of abstract: The Peer Reviews: A Major Lever For Improving Quality

Bonnans Dominique – INSEE, Montrouge FRANCE

The peer reviews set up following the adoption in 2005 of the European Statistics Code of Practice, aim, as a priority, to verify NSIs compliance with the principles of the Code and enable them to benefit from the wise counsel of their peers. In this way, in accordance with the advice following our first review, governance of the French Official Statistical Service was meaningfully reinforced. This review also triggered the implementation of large-scale quality actions, promoting the assimilation of the code by all our agents. Indeed, our second review, which concluded in 2014 that Insee had achieved a high level of compliance, consequently inspired an ambitious Action plan. These actions aimed in particular at integrating quality into our statistical processes and further formalising our quality approaches. As a matter of fact, our third review highlights the significant progress achieved over the last few years in this particular area. In addition, two of the peers' recommendations are actually of high importance for the further development of official statistics. The first one aims to systematically involve Insee and Ministerial Statistical Offices when projects regarding administrative data are likely to have an impact on the production of our official statistics. The second one concerns widening the access of the Official Statistical Service to privately-held data. We are convinced that our third peers' report will be extremely valuable to us when advocating for these two key areas of progress. Beyond this "audit" function, peer reviews also provide an excellent opportunity for us to get all our agents to commit themselves to a sustainable dynamic of continuous improvement, both at Insee and more widely throughout the Official Statistical Offices. This commitment reinforces, in fact, a growing tendency to incorporate Quality in our practices in a sustainable way.

Keywords: peer review; continuous improvement; integrated Quality.

Title of abstract: Peer reviews and modernisation/innovation

Carsten Zangenberg, Karin Blix – Statistics Denmark, Copenhagen, Denmark

The third round of peer reviews in the ESS is being implemented in the years 2021–2023. Denmark will be subject to a peer review visit in March 2022. The process of preparing for the peer review started already in May 2021. The preparation process itself was a fruitful process for the Danish Statistical system. In the preparation, not only a greater awareness of the ES CoP was built up both in the NSI and in the participating ONAs, but activities for developments in the statistical system was also launched. The process served as an encouragement to further developments in the Danish national statistical systems.

This paper discusses the experience from the third round of peer reviews in Denmark and how this will be used in the process of modernising the Danish statistical system. It will discuss our strategy for the coming years and what measures have already been taken for new developments and what is in the pipeline for the near future with focus on standardisation of quality assurance processes and communication.

Keywords: ESS peer reviews, quality assurance, trust, communication.

Title of abstract: Ensuring high-quality data in the new data ecosystem – the 2021 peer review of Eurostat

Aurel Schubert – Vienna University of Economics and Business, Chair, European Statistical Governance Advisory Board (ESGAB)

The third peer review of the European Statistical System (ESS) started in mid-2021 and will cover by mid-2023 all its members. In the summer of 2021, ESGAB performed the peer review of Eurostat. The focus on the observation of the European Statistics Code of Practice had this time a strong future orientation, thus recognising that a solid institutional framework is a necessary but by no means sufficient condition for staying relevant by delivering trustworthy and trusted high-quality statistics in a rapidly changing data ecosystem.

This presentation will discuss the observations and 18 concrete recommendations of ESGAB for Eurostat with a special focus on those that have the goal to enlarge the production possibilities frontier for official European statistics. Among these are the recommendations to develop a comprehensive strategy for the use of new digital data sources, as well as to identify the necessary skills and competences of staff for the future in order to be able to respond to the challenges and opportunities of the new data ecosystem and so be and remain competitive relative to potential private information providers. A legally secured and sustainable access of official statistics to the new sources of privately-held data is a precondition for all this, and so a recommendation of ESGAB directed at the European law makers. Also, an even stronger cooperation including the sharing of confidential data and registers between Eurostat and the European Central Bank's statistics function would help enlarging the output of high-quality European statistics.

Keywords: peer reviews, Eurostat, ESGAB, official statistics.

Session 6
MODERNISATION OF QUALITY FRAMEWORKS

Title of abstract: A new approach to assure quality in Norwegian official statistics

Coen Hendriks, Grete Olsen, Tore Nøtnæs, Frode Berglund – Statistics Norway, Norway

The paper describes a new approach for the evaluation of the quality in Norwegian official statistics. A new approach was needed because a new Statistics law was adopted in 2019. The law has specific quality requirements for official statistics. The requirements are in line with the intentions of the Code of Practice, but not every principle from the Code of Practice is mentioned in the law. The law states also that Statistics Norway shall prepare an annual public report to the Ministry of Finance on the quality of official statistics. Together, the Code of Practice and the statistics law form the framework for quality in Norwegian official statistics.

To meet the requirements from the framework, Statistics Norway is developing a system for quality assurance of official statistics as a whole. This means that official statistics which are produced by other national authorities also must be included in the system. Statistics Norway has a broad experience in assuring the quality of statistical processes in Statistics Norway. Due to the new statistics act, there is a need to collect data on the quality of all official statistics. For this purpose, a new approach was developed.

Experience from using the Code of Practice in internal quality audits shows that colleagues find it difficult to assess whether a statistical process is compliant with a specific indicator. This is because there can be a number of quality dimensions in one indicator. Therefor we applied the principles, indicators and dimensions from the Code of Practice into a user-friendly approach.

The new approach consists of a two-stage collection of data on quality among every responsible producer of Norwegian official statistics. The first stage is a self-administrating questionnaire survey. The second stage is an interview to follow up and clarify information from the survey and to collect additional information.

A questionnaire and an interview guide were developed from the principles and indicators from the Code of Practice, taking the legal requirements into account.

The paper will give a description of the new framework requirements for the production of official statistics in Norway. It will also give a brief description of the other elements in the system for quality assurance of official statistics. The main object of the paper is to give insight into the process of developing the tools for the two-stage data collection and to share experiences from the first round of data collection.

Keywords: quality assurance, Code of Practise, quality questionnaire, national program, official statistics.

Title of abstract: Boosting the National Statistical System through an ad hoc quality assurance framework

Gabriele Ascari, Andrea Bruni, Mariateresa Fiocca, Giorgia Simeoni – Istat, Rome, Italy

Established in 1989, the Italian National Statistical System (Sistan) is the network of public and private bodies providing official statistical information at national and international level. Such a network includes, beside Istat, the Other National Authorities producing European statistics (ONAs) as well as other institutions involved in the production of national official statistics (not-ONAs). It is a significantly complex System since it comprises heterogeneous bodies differing in size, territorial areas, socio-economic contexts, level of government.

As well known, the European Statistical System (ESS) Common Quality Framework applies to Istat and ONAs, while for the not-ONAs an Italian adaptation of the European Statistics Code of Practice (ES CoP) was adopted in 2010.

Yet, since then, several changes occurred in statistics production due to a significant evolution in the Information Technology and Data Science areas, a great exploitation of administrative data and the availability of new data sources, and so on. Accordingly, the ES CoP was revised twice since 2010.

The need for a revision of the national Code was widely recognized. Thus, the new "Italian Code for the Quality of Official Statistics" has been developed by Istat and approved by the Committee for Policy and Coordination of Statistical Information (COMSTAT) in 2021. It includes 16 Principles mirroring the ES CoP ones, and 61 Criteria tailoring the 84 Indicators of the ES CoP.

Still, the new Code alone is not enough to support Sistan not-ONAs quality improvement. A roadmap aimed at introducing a mechanism for continuous quality improvement has been defined.

First of all, as pointed out by the Commission on Quality Assurance of Statistical Information (COGIS) and COMSTAT, two accompanying documents are worth being developed: i) a specific Quality assurance framework, aimed at defining the appropriate methods to implement the Code Principles and Criteria; ii) a national handbook for metadata and quality reporting based on the ESS Handbook for Quality and Metadata Reports.

The required Quality assurance framework is currently being developed at Istat thanks to the joint efforts of researchers working in the Quality team and in the Sistan sector, and it also involves several Sistan not-ONAs as reviewers, in order to pool the work experiences and tailor the methods for the heterogenous producers of official statistics.

Once the documents will be finalised, they should be shared with the Sistan non-ONAs through workshops and capacity development initiatives. Afterwards, a mechanism for the assessment against the Code will be developed and applied, in order to identify best practice and actions to improve the quality level of outputs.

This procedure will increase the compliance with the Code till the next assessment, triggering a virtuous cycle of continuous quality improvement.

The aim of the paper is focusing on the roadmap and the tools that are being developed for its implementation.

Keywords: National Statistical System; Quality Assurance Framework; Roadmap.

Title of abstract: Quality Framework for combining survey, administrative and big data for official statistics

Yvonne Gootzen, Piet Daas, Arnout van Delden – Statistics Netherlands, The Hague, Netherlands

Creating statistics by combining data sources allows for the production of new, more timely and/or more detailed statistics. With a foreseen statistical output in mind, and various potentially useful data sources, a logical first step is to assess the ability of each source to contribute to the foreseen output. Quality frameworks provide tools for such tasks.

Most multi-source statistics are based on a combination of surveys and administrative data. Meanwhile, the first examples of successful applications of big data in official statistics are appearing (ESSnet Big Data, 2018) (Puts, Daas, Tennekes, & de Blois, 2018). Big data comes with its own challenges, which are different from the challenges regarding the use of surveys or administrative data for official statistics. The experience at Statistics Netherlands reveals that frameworks created for multi-source statistics based on surveys and administrative data alone cannot be fully applied to big data sources (Groves & Lyberg, 2010) (Zheng, 2015). In such a framework, a big data source would usually be included in the "other" category, which is very uninformative. Similarly, quality frameworks specifically developed for big data are not designed to include the most relevant information in surveys and administrative data (Essnet Big Data II, 2020). Therefore, the need arises for a multi-source quality framework that is relevant and applicable to surveys, administrative data and big data.

The paper reports on the similarities observed between existing quality frameworks and proposes a framework based on combinations of surveys, administrative data and big data. In the proposed framework, the sources that one is interested in to combine are assessed individually. In a step that will be developed later, the information per source is combined to judge whether the combination of sources be used to produce the intended statistic. In terms of the hyperdimensions of Karr, Sanil, & Banks (2006), the proposed framework mainly focusses on the data dimension of quality. The framework takes into account the target variable and target population of the intended statistic. Additionally, it includes the intended aggregation level of the foreseen statistic and any available auxiliary data.

Various scenarios of combining sources can be identified with the framework. It includes dimensions applicable to survey, administrative and big data in such a way that each of them is informative for the particular foreseen use. Many of the dimensions are also applied in the quality evaluation of statistics based on survey data. The list of dimensions is comprised of Relevance, Accuracy (with sub-dimensions: Population Coverage, Population Representativity, Concept Validity, Concept Stability and Correctability), Timeliness and Punctuality (with sub-dimensions: Recentness and Processing Speed), Accessibility, Meta-data and Comparability. In the paper some applications are discussed.

Keywords: combining data sources; multi-source statistics; data integration; quality framework.

References:

ESSnet Big Data. (2018). Report about possible new statistical output based on (European) AIS data. Eurostat.

Essnet Big Data II. (2020). Work package K: Methodology and quality. Eurostat.

Groves, R. M., & Lyberg, L. (2010). Total survey error. Public Opinion Quarterly, 849-879.

Karr, A. F., Sanil, A. P., & Banks, D. L. (2006). Data quality: A statistical perspective. Statistical Methodology, 137-173.

Puts, M. J., Daas, P. J., Tennekes, M., & de Blois, C. (2018). Using huge amounts of road sensor data for official statistics. AIMS Mathematics, 12-25.

Zheng, Y. (2015). Methodologies for cross-domain data fusion: an overview. IEEE Transactions on big Data, 16-34.

Title of abstract: The role of quality in the core values of official statistics

Steven Vale – UNECE, Geneva, Switzerland

The official statistics community is guided by a set of core values, which inform our everyday activities and are reflected in our standards and practices. They provide a framework for decision-making, and for interactions with governments, society and other stakeholders. They support national and international collaboration and communication by helping to articulate what sets official statistics apart from other data. An international task team under the Conference of European Statisticians has been tasked with elaborating the list of these core values, including descriptions and mappings to other relevant standards, and quality frameworks.

This presentation will provide a summary of the work that the task team has done so far, with a particular focus on how the core values relate to existing quality frameworks. It will highlight the complementary nature of core values and quality concepts, and how they can be used together to communicate the nature and importance of official statistics to the public and other user groups. It will also consider possible next steps and future directions for work in this area.

Keywords: core values; transparency; relevance.

Session 7
CHALLENGES IN QUALITY MANAGEMENT

Title of abstract: Integrating Quality Management and Enterprise Architecture: the Hungarian use case

Józsefné Ferencz, Kriszta Gáti-Kiri, Andrea Petres, Zoltán Vereczkei – Hungarian Central Statistical Office, Budapest, Hungary

Improving the quality of the data, that of statistical products, processes, and the management system are considered a priority for all national statistical organizations. In order to improve operational efficiency and to enable the organization to quickly adapt to the fast-paced changes in its environment and in user requirements, the Hungarian Central Statistical Office aims at constant improvement of its quality management system.

As processes are at the core of the quality management system for any National Statistical Institute, statistical data production and supporting processes had to be examined. We followed an approach that is in accordance with architecture management principles: describing all processes using standardized, reusable objects, building blocks; all based on internationally recognised ModernStats standards (GSBPM, GSIM, GAMSO) and Hungarian national standard solutions and using an IT tool that is recommended for modelling enterprise information. This work results in a powerful tool that contains all the information about processes, its elements and the relationships among them and that is suitable to provide information for different levels of decision-making.

Integrating the use of this tool into the internal decision-making processes supports evidence-based operational solutions and efficient decision-making in various areas: strategic planning, capability development, workflow optimization, methodological standardization, standardization of IT services, HR practices and many more. The paper provides insight into our implemented solution, presents some of the benefits of our approach, lessons learned so far, and the road the Hungarian Central Statistical Office took to build an integrated system that is able to enhance continuous improvement within the organization.

Keywords: quality management, enterprise architecture, standards.

Title of abstract: Conceptualization, design, and implementation of the quality framework for the statistical use of administrative records in Colombia, based on the integration of learnings from NSO experience and the Blue-ETS

Julieth A Solano, Juan Sebastian Ordoñez, Mateo Cardona, Sandra Salamanca – Departamento Administrativo Nacional de Estadística (DANE), Bogotá, Colombia

The use of administrative data (i.e., registers) as source for the production of official statistics has intensified in recent years in many National Statistical Institutes (NSI's). The statistics based on administrative records provide alternatives for the production of high public value information and enable the implementation of integrated systems of statistical records. A correct and efficient use of administrative data require a broad governance ecosystem that allows data to be tracker throughout its entire life cycle inside the NSI. A key part of this governance is the design and application of a procedure to determine the quality of administrative data in a straightforward and standardized way that combines the data source quality approach and the input-oriented output quality approach (Daas and Ossen, 2011), it also integrates recent standards for quality management developed by BLUE-Enterprise and Trade Statistics

From DANE´s experience, the use of administrative data generates challenges in terms of technological architecture, data management and security, due to the higher volume and frequency of the data, and in terms of their quality. To tackle this last challenge, the DANE has designed and implemented strategies for the short and medium term. In the short term, we have implemented agile methodologies such as the Peer Reviews of key administrative registers to generate synergies between the NSS entities responsible for administrative records and their users, develop strategies to ensure quality and encourage consistent statistical use, this strategy allows to recognise the key role of some administrative records as the social security payments and the importance of ensuring coherence of the statistical use of these data from different entities of the NSS. In the medium term, we have designed a program to strengthen administrative registers to work closely with providers in the diagnosis and identification of challenges and improvements needed to facilitate the statistical use, as well as a policy for the governance of administrative registers and alternative data sources. Furthermore, we have developed pilots for the automation of centralized technical checks in the IT office and delivered specific statistical results based on the creation of statistical records with quality management though applications for employment, poverty among others.

Keywords: Administrative records; data governance; quality management, peer reviews, strengthening program, quality hyperdimensions.

Title of abstract: The end of a cycle: the evolution and progress of INE's Quality Management System facing the challenges of the digital data age

Agustín Cañada, Luisa Muñoz, Asunción Piñán, Marcelino García – National Statistics Institute of Spain (INE), Madrid, Spain

Nowadays, National Statistical Offices (NSO) are involved in a process of severe changes which is connected to the data revolution. The access to new digital sources is an opportunity to complement or even replace current sources like the traditional probabilistic surveys, which progressively are suffering from lower response rates and higher collection costs, and the increasingly administrative sources. Following this approach, the Spanish NSI has also integrated not only these data sources (mobile phone data, Scanner data, web scrapping…) but also new methodologies after Eurostat initiatives (such as ESSnet Big Data, in which INE takes part).

The new paradigm and multisource scenario require consequently adapting the Quality Management Systems (QMS): Despite the main elements of the Common Quality Framework in the European Statistics have been updated to meet these data revolution trends (the CoP was updated in 2017 and the QAF in 2019), traditional sources and approaches still conform the core of those models. There is still room for improvement, until those global schemes comply with the specifications and features of new sources and doctrine.

In this scenario, the irruption of COVID-19 has been a major challenge for statistical offices, which have had to introduce, at least temporarily, significant changes in the orientation of their production and in their production processes.

To illustrate the challenges involved in the adaptation of NSO-QMS, in this paper we share our experience in this regard, focusing on five key points:

- Output quality. Besides the quality programmes stated by the European regulations, INE has specific tools designed for the internal evaluation of this variable: a "barometer" of quality indicators for each product; an internal "Quality report" for each department; reviewing of key statistical outputs performed by external experts.

- Standardization of statistical processes: After several implementation waves (starting in 2014) INE statistical operations are documented following the GSBPM standard.

- User satisfaction & needs. Surveys are the traditional method of gathering the overall user feedback on the INE statistics. Since 2007 INE has been conducting this type of surveys to qualified users, once every three years. Simultaneously, some other instruments and procedures to indirectly approach the users are being used (access data to web, other media analysis).

– Finally, from an institutional perspective, it is worth to highlight the central role played by the Quality Committee (collegiate body formed by every INE dpt.) for the improvement and upgrade of the QMS.

Title of abstract: A quality management framework for large-scale multi-source micro data systems

Johannes Kleibl, Johannes Micheler – European Central Bank, Frankfurt am Main, EU

Over the past years, large-scale micro databases on financial market participants and financial instruments such as securities, loans or derivatives have become increasingly relevant for producing official statistics. However, so far there is no common framework for the data quality management of large-scale micro data for official statistics. In this paper, we aim at addressing this gap. The paper starts off with a case study of the data quality management framework of the European System of Central Banks' Centralised Securities Database (CSDB), the largest security-by-security micro database in the central banking world. Using the case of the CSDB, it identifies several challenges in conducting data quality management for large volumes of micro data that are regularly reported and updated from multiple data sources that are overlapping and potentially inconsistent.

To address these challenges, the paper derives a set of core principles for the efficient and effective data quality management of large-scale multi-source micro data systems. These principles underscore the importance of (a) the standardisation of micro data, (b) the integration of data quality management workflows, (c) the use of automated and targeted rules for identifying data quality issues, and (d) the prioritisation of verification work. Based on these principles, the paper proposes a unified "360-degree framework" for the data quality management in large-scale micro data systems. In doing so, it makes a first step towards developing a general framework for the quality management of large-scale micro data for official statistics.

Title of abstract: Could there be Guidelines for developing and maintaining quality guidelines?

Thomas Burg – Statistics Austria, Vienna, Austria

Regarding principle 4 of the Code of Practice for European Statistics the existence of quality guidelines represent an essential method in order to communicate to producers and users that official statistics has to be produced according to well specified standards. Most NISs have made available some sort of Quality Guidelines – for instance some of them reacting on a peer review recommendation – generically relevant for most important process steps. But in the last couple of years triggered by new developments quality guidelines were or are currently produced covering more specifics topic (use of new sources, geo-statistics, use of frames etc.).

The paper starts by showing comparing some examples of existing quality guidelines to show in how far they differ in structure and contents. Followed by that it is discussed if there can be defined certain elements which could be identified as relevant for quality guidelines. Questions related for that are

– In which way should quality guidelines be presented?
– Is it necessary to provide minimum requirements?
– How is he relation to guidelines and the process model in use (for instance GSBPM)?
– How detailed should the question of quality measurement be addressed?

Since we are living in an accelerated world where the relevance of quality guidelines is also dependent in how far it is possible to ensure the actuality of the guidelines regarding new developments the paper discusses finally possible maintenance procedures for quality guidelines.

Session 8
QUALITY MANAGEMENT SYSTEMS

Title of abstract: Promoting quality framework in the National Statistical System and beyond

Jolanta Minkevica – Central Statistical Bureau of Latvia, Riga, Latvia

The European Statistics Code of Practice (the Code) is the cornerstone of the common quality framework of the European Statistical System. It is a self-regulatory instrument and is based on 16 Principles covering the institutional environment, statistical processes and statistical outputs. Nevertheless, in the institution which main business is to produce official statistics, it is not enough. This is where the quality management system (QMS) comes in.

Agile maintenance of the QMS in the National Statistical Institution (NSI) as well as coordination of the implementation of requirements of the Code in the National Statistical System (NSS), done by the Central Statistical Bureau of Latvia (CSB), aims to ensure that statistics produced are relevant, timely and accurate and to meet the needs of our customers and other stakeholders more effectively.

Both – the CSB QMS compliance with the standard (ISO 9001) and the NSS compliance with the Code – are periodically assessed internally and by the third parties. The presentation will provide an insight into main aspects and the steps taken by the CSB so to promote implementation of quality management in the NSI and the NSS and achieved results.

Keywords: Code of Practice; quality framework; coordination.

Title of abstract: Quality management in the Czech statistical office

Petra Kuncová, Martina Górska – Czech Statistical Office, Prague, Czech Republic

The Czech Statistical Office forms part of the public administration of the Czech Republic and, therefore, it must also meet the requirements of the Government of the Czech Republic. In recent years, activities related to the introduction of quality into state and public administration offices have become more significant. Within this context, the Government of the Czech Republic has adopted a resolution requiring all authorities to introduce quality management elements, either to a minimum extent, as represented by ten improvement criteria, or to an optimum degree, fulfilled by obtaining an ISO 9001 certificate. The project is managed by the Ministry of the Interior of the Czech Republic.

The CZSO was one of the top quality management authorities in 2006–2009. It even won a national quality award for self-assessment with the EFQM model. After ten years, when activities in the area were not supported in the CZSO, a great deal of emphasis was once again placed on them. The project of the Ministry of the Interior of the Czech Republic described above has become an opportunity to start

developing activities systematically again. To implement quality at a minimum level, it set out 10 improvement criteria (1–10):

1. Office Development Strategy
2. System for Identifying Measurable Objectives of the Office
3. Role, Structure and Powers of the Office
4. Internal Regulations System
5. Office Communication System
6. Change Management System
7. Human Resources Policy
8. Adaptation Process
9. Employee Satisfaction Survey
10. Civil Service Partnership System

Mandatory outputs are set for each criterion but compliance with the criteria can also be extended in accordance with the needs of the Authority.

The CZSO has received financial support from the European structural funds for this project. We are now in the process of implementing the project, whose fulfilment has been influenced to a large extent by the Covid pandemic, especially in the field of organising different training courses. Completion of the project is scheduled for 31/12/2022. In the presentation we want to report on the status of the project and its links to statistical activities, especially a peer review.

Title of abstract: The contribution of quality management system requirements for quality assurance in official statistical production

Rogério Reis, David Sousa – Statistics Portugal, Lisbon, Portugal

Quality management systems in statistics are an active area of research and debate both at national and international level. In fact, the performance of national statistical institutes (NSI) is affected by quality management practices. Recent developments in statistical production and dissemination, the establishment of codes of best practices for national statistical systems, as well the development and approval of international standards and quality tools, among others, provide some examples.

The importance of the European Statistics Code of Practice (CoP) and International Standardization Organization (ISO) standards for quality assessment and reporting, as well the inherent debate on how well quality requirements are fulfilled, has become crucial for the renovation of a quality policy in order to adapt to the new production environment. Furthermore, the outcome of the debate is important for the innovation objectives fostered by the top management of the Portuguese NSI.

This paper examines the volume of the debate on key issues raised on quality evaluation and monitorisation actions, taking stock of areas like conceptual metadata, document management, skills and competences, processes, organisation, structure, and resources. In addition, this paper presents ideas and an analysis about the role

that an ISO 9001:2015 certification can take in the fulfilment of quality requirements and in improvement actions regarding CoP compliance, for an innovative, sustainable, and high-quality performance.

The analysis shows that capabilities in improvement, innovation, sensing weak signals, and responsiveness, all help sustain high-quality performance. Furthermore, this paper presents evidence that an empirical causal relationship between quality performance and innovation performance can be found, suggesting that the achievement of one aspect of performance has an impact on the other.

Keywords: quality requirements; quality assurance; ISO 9001:2015; European Statistics Code of Practice.

Title of abstract: A new approach to monitoring the quality features of projects in the Italian statistical programme

Mauro Scanu, Maria Pia Sorvillo – Istat – Rome – Italy

The Italian statistical system (Sistan) is a large network of statistical offices located in Ministries and other central institutions, territorial authorities, and some private organisations. They amount to more than 3 thousand, but only a small part is involved in the production of official statistics at the national level, contributing to the National statistical programme (Psn). For this small group, including 58 entities, it is especially relevant to deepen their approach to quality.

The main idea is to rely, also at a local level, on the directions and approaches suggested at a more general European level through the European statistical code of practice (European Statistical System, 2019) and the ESS Quality assurance framework (Eurostat and European Statistical System, 2017). For instance, a Quality assurance framework for Sistan (non-Other National Authorities, ONA) agencies is currently under development in Italy.

Under this framework, for the first time the commitment of Sistan authorities to quality has been analysed looking at answers in a section of the questionnaire filled in by each entity for each project included in the Psn. Three topics are covered: metadata; variables' definitions and classifications; data sources validation.

Considering all the Sistan projects included in the 2021 Psn (more than 700), most of them take into account metadata for informative contents (78%) whereas the percentage decreases to 46% for metadata about the process and to 40% for quality indicators.

Official classifications are used in almost all projects (86%) – especially for territorial units – while less than 50% refer to official variables' definitions.

A large part of projects relying on administrative data sources foresee a validation step (83%), in most cases in terms of a check on completeness and internal consistence, while an assessment of classifications/definitions and a cross validation, using other data sources, are more rarely performed.

A deeper analysis by single entity or by project category, as well as a revision of the questionnaire to better specify quality features of the Psn projects, can help identify weaker areas that need more attention in terms of awareness raising and training. This is particularly relevant as Sistan will be most probably asked to contribute effectively to monitoring of the National plan for recovery and resilience, funded by EU in the frame of the Next generation program.

Bibliography

European Statistical System (2019). *European statistical code of practice*, Version 2.0 URL https://ec.europa.eu/eurostat/web/quality/european-quality-standards/quality-assurance-framework

Eurostat and European Statistical System (2017). *ESS Quality assurance framework for the National Authorities and Eurostat (EU Statistical Authority)*. URL https://ec.europa.eu/eurostat/web/products-catalogues/-/KS-02-18-142

Keywords: national statistical programme; quality monitoring; harmonisation.

Title of abstract: Quality Management System and Documentation of the Quality of Statistical Surveys

Lidija Brković – Croatian Bureau of Statistics, Zagreb, Croatia

The Croatian Bureau of Statistics has presented the results of the project "Quality Management System and Documentation of the Quality of Statistical Surveys", which has improved the quality management system to reach the level set by Eurostat's strategic document The ESS Vision 2020 and the ESS Quality Assurance Framework implementation document.

The budget of EUR 240.000 withdrawn from the European Structural and Investment Funds has been invested in upgrading the functionality for monitoring the quality of statistical products in the POMI application and database, thus fully harmonising the quality management system of the Croatian Bureau of Statistics with the quality reporting standard prescribed by Eurostat.

Over the last four years of the project, new tools have been added to monitor the quality of statistical products, which has further strengthened the capacity of the Croatian Bureau of Statistics and ensured even better data quality. By upgrading the POMI application and database with information on the quality of realised statistical products, the Croatian Bureau of Statistics has once again confirmed its focus on the quality of all processes, products and services, following the quality principles defined by the European Statistics Code of Practice, namely the relevance, accuracy and reliability, timeliness and respect for publication deadlines, consistency and comparability, and availability and clarity of data.

Information on the quality of statistical products is available to users in quality reports, which are a kind of a certificate guaranteeing that statistical data are collected, processed and disseminated in accordance with European standards. Quality reports are available on the website of the Croatian Bureau of Statistics.

Keywords: quality management system; quality reports.

Session 9
COORDINATION AND COOPERATION IN THE NSI, NSS AND BEYOND

Title of abstract: Peer reviews of ONAs as part of the coordination and cooperation in the NSS

Karin Blix – Statistics Denmark, Copenhagen, Denmark

As part of the implementation of the new regulation for European statistics, Statistics Denmark has developed a set of guidelines for European statistics for the Other National Authorities (ONAs) producing European statistics. The guidelines are based on the ESCoP, but adapted to authorities that do not have statistics production as its core business.

The development of the guidelines started in 2015. The ONAs were consulted when drafting the first version of the guidelines, and a self-assessment questionnaire was developed and used in the first round of monitoring compliance in 2016. This round made it clear that the ONA's had different interpretations of the guidelines. Statistics Denmark decided to visit all the ONA's to better understand their production systems and challenges. In these visits, information was gathered to update the guidelines for the next round of monitoring compliance. The guidelines have been adjusted every year since, to ensure better comprehension. Compliance with the guidelines is monitored annually by a self-assessment questionnaire and approximately every five years by a more thorough peer review of the ONAs. The peer reviews are carried out by Statistics Denmark with the participation of two representatives from other ONAs.

This paper will discuss how the peer reviews are carried out and how they contribute to improving the quality in the production of official statistics and how it contributes to better coordination and cooperation in the NSS.

Keywords: Coordination, Cooperation, Guidelines, Monitoring, Peer Review.

Title of abstract: Official statistics methodology network as a tool in the Norwegian quality framework

Kari-Anne Lund, Xeni Dimakos – Statistics Norway, Oslo, Norway

The new statistics act has given Statistics Norway a double role as both a producer of official statistics, and leader of the Committee for official statistics. The statistics act states that Statistics Norway should coordinate official statistics and report on quality. However, each producer is responsible for the quality of their statistics. This is the backdrop for Statistics Norway to take on the responsibility for supporting other producers of official statistics, but without taking responsibility for the quality of all

official statistics. This paper for Q2022 describes the establishment of the new methodology network for producers of official statistics. Statistics Norway has suggested that establishing such a network for methods would be an efficient means to the goal of quality control. The network should encourage competence building among and across member of the committee, as a tool to brand "official statistics".

The Committee for official statistics was appointed by the Norwegian Ministry of Finance in 2019. The committee shall contribute to an expedient and effective national statistical system. Of the 24 institutions who are contributing in the committee, 12 are producers of official statistics. Statistics Norway has suggested several components in a system for quality assurance of official statistics. One of the elements suggested was to establish a methodology network for the participating institutions. This was motivated by a desire to create a meeting place that encourages cooperation and exchange of experiences.

The overall purpose of the methodology network is to contribute to improved quality of official statistics. This is achieved by encouraging cooperation on statistical methods across institutions who produce official statistics. The methodology network serves as a forum for professional exchange, discussions, information sharing and competence building.

This paper describes the establishment of the new network. It also gives a description of how one may seek to explore which are the most crucial methodology topics to establish a basis for mutual exchange of knowledge. Further, it describes how the network is run and monitored in practice. In continuance of these organizational descriptions, we discuss our strategy to use this new methodology network as a tool to confront the brand "official statistics". Among other questions to enlighten in the paper, are:

- What is the role and responsibility of Statistics Norway and each network member?
- What are the network costs and benefits for Statistics Norway and the other members?
- How is the experiences and response from the first year?
- How was the network established?
- What are the future plans for the network?

The main object of the paper is to contribute to a shared understanding of how methodology networks may be established and aimfully used to enhance quality of official statistics.

Keywords: quality frameworks, coordination and cooperation, official statistics.

Title of abstract: Challenges of coordination and cooperation in the production of SDG indicators in Spain

Pedro Revilla, Antonio Salcedo, Ana Carmen Saura – INE, Madrid, Spain

A global commitment to development and transformation such as the 2030 Agenda, made up of measurable targets, requires quality statistical information. The National Statistical Institute of Spain (INE) is committed to the production of high-quality indicators that ensure an appropriate monitoring and reporting of the goals and targets of the 2030 Agenda. In all its actions, INE is based on the principles of the EU Code of Practice as well as on the collection of methods and tools of the EU Quality Assurance Framework. The full development of the global indicator framework, complemented by regional, national, sub-national and thematic indicators, presents an unprecedented challenge for statistical systems. In the case of Spain, there is an added difficulty since it has a decentralized system, both departmentally and territorially. An example of this decentralization is that, although INE is the institution of the Spanish Statistical System that individually produces more indicators, the Other National Authorities (ONAs) jointly produce more indicators than INE.

The complex system of production of SDG indicators in Spain, characterized by territorial and departmental dispersion, presents many coordination and cooperation challenges. The set of indicators of best practices and standards of the recently established Principle 1bis, of Cooperation and coordination, provides a solid frame of reference, which are a great help in articulating a coherent system of information. The United Nations also provides guidelines highlighting the coordinating role of the National Statistical Institutes (NSIs) within the SDGs. Resolution A/RES/71/313 highlights in its paragraph 6 the role of NSIs as coordinators of the national statistical system. UNECE Road Map recommends that NSIs serve as focal points in the measurement of the SDGs.

To coordinate the ONAs responsible for the development, production and dissemination of SDG Indicators, INE has used the legal and organizational tools existing in Spain. A statistical operation called "Indicators of the 2030 Agenda for Sustainable Development" was incorporated into the National Statistical Plan. This operation is coordinated by the Office of the President of INE, in cooperation with the ONAs responsible for the production of the various indicators. The operation was submitted for opinion to the High Council on Statistics, an advisory body where trade unions, business organisations and other social, economic and academic groups are represented, which agreed to unanimously give a favourable opinion. In its role as coordinator, INE has reported and discussed the developments of SDG indicators in the collegiate bodies of coordination (the Interministerial Statistics Commission and the Interterritorial Statistics Committee). In both bodies, Working Groups have been created to study and implement the SDGs indicators, exchange methodological experiences, and solve practical problems.

As a result of coordination and cooperation efforts, it has been possible to build a National Reporting Platform, centralizing all the available data on SDGs for Spain and allowing information to be consulted through a single access point. Currently, the Platform has more than 150 indicators. In addition to INE, 17 ministerial departments and the Bank of Spain participate.

Keywords: Code of Practice, Quality Assurance Framework, SDGs, 2030 Agenda.

Title of abstract: Coordination and the role of quality and standards in the implementation of the Sustainable Development Goals (SDG)

Mogens Grosen Nielsen, Lars Thygesen – Thygesen Statistics Consulting, Copenhagen, Denmark

There is an urgent need for common and reliable information for handling the corona pandemic and climate change, both at national and global level. The work on implementing the SDGs has provided a way of handling these global challenges by introducing common global goals, targets and indicators and methodology.

Compared to traditional surveys the production of SDG indicators often requires new production processes including collection of administrative data from sources outside the national statistical organisation. The organisations involved often use different concepts, have different focus, different values, traditions and methods. This can cause problems when there is a need for common understanding of content and production of indicators. UN, Paris21 and The World Bank has done a lot to address these problems.

However, it is the experience of the authors that many countries are facing huge problems on coordination, aiming at a coherent system with smoothly running production of SDG indicators. The challenge is especially on coordination of work processes where indicators are based on administrative data from data providers outside the national statistical organisations. Many countries find it difficult to get an overview of quality frameworks and standards and how to use and implement these in practice.

This paper claims that there is a need for a more coherent and holistic approach on coordination using quality frameworks and standards to support a common understanding of the implementation of production processes related to SDGs.

Coordination will be discussed in general with focus on communication. Subsequently the paper will look at coordination in two respects: how to ensure that we produce the right statistics based on user needs (nationally and internationally), and how to ensure coordination inside the national statistical system. The paper will introduce quality frameworks, the Generic Statistical Business Process Model and other models and standards.

Thereafter the paper will present

– A methodology for handling changes including description and existing and future solutions to be discussed with all stakeholders. One-size-fits-all does not work, as each country and organisation has its own history and a complex of stakeholders involved in the process.

– A template for a handbook for production of SDG indicators. The template will guide on how to turn principles from quality frameworks and various standards into practice. The handbook must be completed for each country, taking stakeholders' viewpoints into account.

– A simple tool to support the coordination of the production of SDG. The tool will, among other things, include description of each SDG indicator with related metadata, related dataflows and data providers and release calendar for the indicator and dataflows. The descriptions of indicators and dataflows will use the recommended standards for date exchange.

Keywords: coordination; communication; SDG; quality.

Title of abstract: The Macroeconomic Imbalance Procedure (MIP) quality framework and the Eurostat/ECB Memorandum of understanding: improvement of data quality through producers' cooperation

Rosa Ruggeri Cannata – European Commission, Eurostat
Andreas Hertkorn – European Central Bank

Statisticians are committed to supply policy makers with data of highest quality, which is essential for policy decisions at the European level. Eurostat provides indicators for the Macroeconomic Imbalances Procedure (MIP) on the basis of statistics compiled in the Member States either by NSIs or by NCBs.

Close cooperation between Eurostat and the Directorate General Statistics of the ECB, as well as between the European Statistical System (ESS) and the European System of Central Banks (ESCB) is key in ensuring that the quality of European statistics serves the objectives and policies of the European Union. In 2016, Eurostat and the ECB/Directorate General Statistics signed a memorandum of understanding on the quality assurance of statistics underlying the MIP covering statistics in Balance of payments and international investment position and Financial accounts. After five years of implementation, a first balance on achieved results can be drafted. A number of concrete results have been achieved, cooperation between national producers, in particular national statistical institutes and central banks, and between Eurostat and the ECB has been fostered, consistency across domains improved, quality of MIP indicators constantly monitored.

Significant improvements have been noted in the measurement of households' holdings of assets abroad for some member states as well as a better understanding of the coverage and data quality on non-bank financial entities. The quality assessment of financial accounts has been focusing on the reduction of vertical discrepancies between financial and non-financial sector accounts, the coverage of financial corporations and the reduction of discrepancies between annual and quarterly financial accounts. The quality monitoring is supported by MIP visits, where Eurostat and ECB representatives can meet national data producer in order to discuss at a very practical level how to further improve data quality.

The work in improving consistency across domains and data frequencies will continue, based on a very fruitful institutional cooperation among national and supra national institutions, which have been able to handle the challenge of a jointly quality monitoring.

Keywords: quality management; policy indicators; institutional cooperation.

Title of abstract: Tackling data quality issues in UK equality statistics

Rohan Allen, Nadyne Dunkley, Kaite Emmerson, Darren Stillwell, Chris White – Cabinet Office Equality Hub, London, UK

The Cabinet Office Equality Hub focuses on disability policy, ethnic disparities, gender equality, LGBT rights, social mobility and the overall framework of equality legislation for the UK. A central priority of the Equality Hub is improving the quality of data about disparities between protected characteristic groups. High quality government data about outcomes for different groups are vital to inform effective policy interventions to reduce existing disparities.

The Equality Hub has recognised that many data quality issues are shared between different protected characteristic groups. These include the harmonisation of categories; high levels of missing data; and the underrepresentation or exclusion of specific population groups due to sampling design. These quality issues limit the suitability of data to reliably inform policy making. The Equality Hub has formed a Data Quality Working Group comprising of analysts who are experts in the quality of data on ethnicity, sex and gender, sexual orientation, disability and socio-economic background, to address cross-cutting data quality issues. This collaborative approach allows the Equality Hub to address these data issues in a consistent manner, facilitate better sharing of expertise and best practice, and ensure a joined-up approach to data quality to be taken with users and other stakeholders.

The Equality Hub does not collect its own data very often, so improvements come through developing and overseeing programmes of data quality work; providing leadership and guidance to analysts in other UK Departments collecting and analysing data for different characteristics; and collaboration with the Office for National Statistics.

The Equality Hub has launched programmes to improve the quality of the data and evidence for different characteristics including the Ethnicity Quality Improvement Plan; programmes to address COVID-19 health inequalities; a series of Methods and Quality reports; and other improvement work that will also help meet the recommendations of the Inclusive Data Taskforce published in September 2021.

This session describes cross-cutting data quality issues identified by the Equality Hub, the formation and value of the Data Quality Working Group projects, and the implementation and impact of key data quality improvement programmes.

Keywords: data quality, equality, disability, ethnicity, gender, LGBT.

Session 10
GOVERNANCE FOR OFFICIAL STATISTICS

Title of abstract: Managing our core data assets – Practical Data Governance in the Central Statistics Office, Ireland

Ken Moore, Kevin Phelan – Central Statistics Office, Ireland

As the volume, range of sources and access to data continues to increase at pace, it is critical that our core data assets are properly governed so that we can continue to utilise this data to its full potential. Recognising that data governance is central to meeting the complex and ever-changing needs of a National Statistics Institute, the Quality division of the Central Statistics Office, Ireland commenced work in 2018 in developing a data governance framework to meet these needs. This work was also motivated by the evolving data protection legislative environment both nationally and internationally. Good data quality requires effective and practical data management supports and tools and this paper will tell the story of our progress to date.

When developing our plans, we had a number of data related challenges that needed to be resolved. These included a lack of clarity on data ownership and responsibility, a lack of effective governance on data locations and naming standards, some issues regarding access control and data retention and different approaches to metadata standards used to describe the data. When designing our implementation plan, we took the approach from a hands-on, data user perspective rather than the traditional technology driven solution (but obviously collaborating closely with our technology team). For the plan to work the Quality team felt that it had to be driven by a team who understood data, how it was used and the everyday data challenges being faced in statistical production.

The overall aim of the data governance framework was that all statistical data was stored to an agreed standard so that our key corporate asset was better managed and governed. This was based on five key principles: 1. Data is accessible and findable, 2. Data is reusable, 3. Data is supported by relevant metadata, 4. Data has an individual owner and 5. Data is stored in a standardised structure with common naming conventions.

This paper will detail the progress made to date on implementing our data governance framework and describe the tools and structures used. These include our models of data ownership, data structures, data retention and version management. It will also outline our approach to standardised metadata management, the improved governance of access to data and the development of a data management policy to support the implementation of the framework. While there has been good progress made to date, it has not always gone smoothly so we will also detail some of the challenges encountered to date and reference some future plans to further mature our data governance environment.

Keywords: Data Governance, Data Quality, Data Inventory.

Title of abstract: Data ecosystems and NSIs: toward an evolution of institutional and quality frameworks in the production of official statistics?

Stefano Menghinello, Claudio Ceccarelli, Silvano Vitaletti – Italian National Institute of Statistics, Rome, Italy

Data ecosystems are collaborative network of data stakeholders that share and extract value from the integration of multiple data sources. The aim of this paper is to highlight how the institutional and data quality frameworks traditionally adopted by NSIs for the production of official statistics should evolve in order to successfully exploit the wealth of new information included in data ecosystems. In particular, they should shift from an approach based upon a register-based fully integrated but close system of data production to a more open and collaborative data production system where potential threats and opportunities to jointly manage data or sharing knowledge with external stakeholders are carefully manage by NSIs.

An in-depth understanding on how data ecosystems work in terms of stakeholders' specific goals and incentives should represent the starting point of any attempt to establishing a fruitful cooperation between data ecosystems and NSIs. As an example, the stakeholders of a given data ecosystem may have different goals in exploiting the informative value of integrated data, other than official statistics. In addition, their commitment to adopt official classifications, to use methodological standards or to invest a relevant number of human and financial resources in improving the quality of data, may be substantially different with respect to NSI standards. Thus, the new framework of data quality shall evolve to acknowledge the presence of multiple goals in the use of integrated and shared data. It should also define and identify common and diverging drivers and standards in data quality management. In addition, the effective contribution made by an NSI with respect to the data ecosystem may vary widely.

The sharing of data may concern aggregated or micro data, while knowledge transfer from NSI may concern methodology, processes and applications to better manage the classification, integration and data quality assessment, including personal data anonymization or pseudo anonymization techniques. It is crucial to make the contribution made by an NSI to a given ecosystem clear and transparent both internally than externally to the ecosystem network. Ecosystem'stakeholders need to correctly evaluate the level of cooperation of the NSI within the ecosystem. Third parties external to the ecosystem, private companies or institutions, need to verify the integrity of NSI institutional goas and to potential violation of data protection and open competition regulations. In conclusion, this paper will show how the establishment of an effective and positive cooperation between data ecosystems and NSIs relies upon a learning process that should be verified step by step against a new data quality and institutional frameworks that should be capable to well balance not exclusive information benefits with potential legal and data security threats.

Keywords: data ecosystem; quality framework; General Data Protection Regulation (GDPR).

Title of abstract: Building on solid foundations. Lessons on the role of the institutional framework in the quality of official statistics

Hernán Daniel Muñoz – CEPEI. Former National Director of Planning, Institutional and International Relations of INDEC Argentina
Julien Dupont – OECD Statistics and Data Directorate

What key principles should be included in statistical legislations in order to ensure professional independence of the producers of official statistics? How to assess the professional independence of statistical authorities, taking into account the key role of users including the government administration? How to assess the legal basis for the functioning of the national statistical system (NSS), and the production of high-quality official statistics? How to evaluate the authority of the producers of official statistics to provide the reasonable assurance that they are produced and disseminated in accordance with the principles of an efficient quality assurance framework? Based on lessons learned from several relevant case studies, including the situation of the Argentinian statistical system between 2007 and 2019, this paper intends to point out some essential elements required to assess the professional independence of producers of official statistics. The lessons learned from these case studies enable to identify common patterns regarding the *de facto* and *de jure* protections to professional independence required to produce high-quality official statistics.

Keywords: official statistics, modernizing statistical legislation, capacity development, institutional reform, quality.

Title of abstract: The evaluation of the governance system through the Internal Audit System for the production of official statistics

Stamatios Theocharis, PhD – Statistical Head & Head of Department of Internal Audit Methodology & Design, Ministry of Interior, Athens, Greece

Governance is generally recognized as a crucial aspect of the most democratic and effective implementation of public policies through which the parties of the system have a more substantial roll. Public administration through the systems of governance consults and is accountable to the citizens both for the type of public goods as well as for the quality and the conditions of their production. An important parameter of the system is the Internal Control System that consists of a set of functions, procedures and controls that the operator adopts. It is designed to provide reasonable assurance to the entity that it is achieving its objectives, regarding the effectiveness and efficiency of its operations and for the reliability of its reports. The production and dissemination of official statistics is included in the scope of monitoring and control of such a system.

In the field of production of official statistics, the governance structures that have been established by the existing institutional framework in Greece, refer to the combination of structures, procedures and methods applied by the National Statistical Authority in collaboration with the Statistical Heads of public bodies. This system aims at the proper administration and monitoring of the activities of public bodies related to the production of official statistics in accordance with the provisions of Community and

national legislation. This structures also provide an organizational and administrative framework for the relevant services, taking into account the needs of the users, to contribute effectively to the achievement of the objectives of the organization in terms of quality improvement and further dissemination and reuse of the statistics.

The purpose of this work is: a) to investigate the existing governance structure, b) to map the procedures related to the production of official statistics by the Ministry of Interior, c) to identify any risks that threaten the effective production of official statistics and d) to make proposals that will help to ensure the improvement of the quality and the efficient production of statistics. The opportunities provided through the interaction of the bodies with the users for the production of new statistical products will be explored as well.

In the context of the implementation of the provisions for the production of official statistics, the improvement of the quality of statistics and their certification as official statistics in the Ministry of Interior, a Statistical Head has been appointed as well as a relevant Working Group. This structure also seeks to develop a network of two-way communication between the parts of the system as well as the horizontal and uniform application of directives and observations regarding the application of the principles of the Code of Good Practice for European Statistics. In this work we analyse how this structure works, the potential risks and opportunities associated with improving the quality of statistics.

Keywords: Governance System; Risk Assessment; Control System; Audit System.

Title of abstract: Data governance and quality: involving data sources' owners

Bruno Tissot – Bank for International Settlements, Basel, Switzerland
Irena Krizman

An all-encompassing approach to governance is needed when collecting, managing, disseminating and using official statistics and support data quality in a broad sense. Governance frameworks should be holistic, i.e. cover entire organisations including the related principles, policies/procedures, structures, roles and responsibilities, and be an integral part of their strategic plans. But there is also a need to complement any institutional-level approach to data governance by a broader focus covering the entire production and use of national statistics, including alternative sources. This calls for ensuring that private data providers follow adequate codes of principles, clarifying the various responsibilities in the national governance scheme, and establishing proper international guidelines and cooperation mechanisms – as indeed highlighted in the context of the new Data Gaps Initiative envisaged by the G20. Based on observed good practices, especially in Slovenia but also among various IFC jurisdictions, this paper will focus on how to enhance the overall relationship between governance and data quality in the NSS by (i) making owners of information sources aware of the need to strengthen the quality of their data before they can be used; (ii) discussing the roles that official statisticians could play in this endeavour; and (iii) identifying ways to organize the related partnerships and processes.

Keywords: data governance; frameworks; partnerships.

Session 11
QUALITY ASSURANCE

Title of abstract: ES CoP internal compliance review and its positive side-effects

Tina Steenvoorden – Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia

The European Statistics Code of Practice (ES CoP) is nowadays widespread and present in everyday life of European statisticians. Everyone has heard of it and has seen the materials hanging around in the offices. However, how many have actively dealt with the principles and the content behind them? How to address the challenge that everyone has heard a lot about the ES CoP, but not so many have deeper understanding of the principles? How to motivate and include survey statisticians and not only the management to discuss and give active proposals?

The Statistical Office of the Republic of Slovenia (SURS) in 2019 decided to make an exhaustive internal review of the compliance with the ES CoP. The aim of the review was to examine the current implementation of the ES CoP, to gather the proposals for further enhancements from a wide range of employees, and to prepare the action plan for short- and long-term improvements.

The top management gave a strong support to the action and the internal task force was set up to coordinate the internal review and prepare the needed materials. In addition to the management, the rest of the participants were included on a voluntarily basis and were randomly divided into working groups devoted to each principle. Each group prepared from their point of view the description of the compliance and an assessment for the selected principle.

The review turned out to be a success and the positive results have been observed on multiple layers.

–   Based on the materials prepared by the working groups, the internal task force analysed the results, reviewed the assessments and compiled an internal action plan for further improvements.

–   The description of each principle served as a basis for preparing the Internal Handbook on the Implementation of the ES CoP at SURS. The handbook presents for the first time in one place the meaning and practices behind each ES CoP indicator and is a benefit for all employees, but even more for the newcomers.

–   The most important and desired side-effect of the internal review was that the broad scope of the employees substantially increased the awareness of and knowledge about the ES CoP. They have been for the first time included in the discussions about the compliance with the principles, have had a chance to actively search for practices within the office and could give proposals for further improvement actions.

The paper starts by presenting the motivation and the organization of the internal review, followed by the results, plans as well as the description of challenges that we have faced along the way.

Keywords: internal review, European Statistics Code of Practice, internal handbook, improvement plan.

Title of abstract: From audit to assurance – an approach to the delivery of statistical quality assurance and improvement in the CSO

Don Forde, Ken Moore – Central Statistics Office, Cork, Ireland

In 2019, the Central Statistics Office (CSO) piloted a new approach to address gaps in the internal statistical quality assurance environment. This new approach, titled Supported Quality Appraisal (SQA) aimed to go beyond a traditional audit focus on compliance towards the delivery of continuous quality improvement. This paper describes CSO's experience with this new system for statistical quality assurance and its goal of building an internal quality assurance structure that not only identifies quality issues but also ensures the delivery of real change in statistical production.

The CSO SQA model, involves bringing together multi-disciplinary expertise from Quality, Methodology and Technology areas to work with statistical production areas to review and assess current practice. This team takes an independent look at existing statistical production by examining documentation, quality information, methodologies and production systems. Processes are reviewed in discussions with the statistical business area. The SQA team identifies good practices in the area that might benefit other statistical production areas and explores options for improvement of existing practices.

Appraisal findings are agreed with the statistical business area. Together, the independent appraisal team and the business area then work out action plans to deliver improvements in a practical, prioritised manner – the supports required, the sequencing and scheduling of improvement actions are agreed, and both the business area and appraisal team take an approach of joint responsibility to implementation of improvements. Quality Unit coordinates the work, follows-up to ensure actions are implemented and communicates key findings, good practices and lessons learned to senior management and to the wider office.

A key benefit of the SQA approach is how it gains crucial buy-in from statistical business areas. It does this by creating a sense of "team", built on recognition that perfection only exists as an aspiration, that improvements can always be found and by giving a commitment to hands-on support in delivering recommended changes. Because of this emphasis on identifying opportunities for improvement as a team, SQAs avoid drifting into a narrow fault finding or "blame game" exercise. This builds trust and gains the buy-in that is critical in implementation of change. A fundamental

principle of the model is that the focus is on looking forward to the delivery of continuous quality improvement.

This paper will give an overview of the approach taken in the three Supported Quality Appraisals undertaken in the CSO and will describe how these have worked, some difficulties that were encountered, including the improvements that have been made, the impact of the Covid-19 situation, and the key lessons learned over the past two years.

Keywords: statistical quality assurance, multidisciplinary appraisal, collaborative approach, continuous improvement.

Title of abstract: Managing risks to statistical quality in the UK Office for National Statistics

Charles Lound, David Mais – Office for National Statistics, United Kingdom

The UK Office for National Statistics (ONS) is implementing a Statistical Quality Improvement Strategy that sets out our renewed organisational commitment to quality and the actions we will take to improve the quality of our statistics and analysis. The Strategy includes five strategic objectives for improving statistical quality. The ONS Quality Committee was recently established to oversee the strategic direction and provide assurance over the quality of statistics.

To help provide this assurance, the Quality Committee has identified a series of lines of defence. These include placing responsibility for the quality of individual outputs with the Division that produces that output. They also include a Statistical Quality Maturity Model (SQMM) and a series of deep dive quality reviews, together with Divisional Quality Improvement Plans and a network of Quality Champions. Further lines of defence come from Audit and Risk within ONS working closely with the Quality Committee on statistical quality, and the Office for Statistics Regulation that has the formal role in assessing compliance with the UK Code of Practice for Statistics. In the Code, quality means that statistics fit their intended uses, are based on appropriate data and methods, and are not materially misleading.

In the SQMM, producers score their division and each of their outputs on a series of questions relating to quality, justifying each response, leading to an overall score. In the deep dive quality reviews, we examine the end-to-end production process, looking for risks to quality. These lead to recommendations to address these risks, with support from the Methodology and Quality Directorate and others.

In this paper we will describe the first year of implementation of the strategy, including the outcomes of the SQMM and deep dive exercises, the way the lines of defence have been adapted to have the best likely impact on risk to statistical quality.

Keywords: statistical; quality; risk; institutional; management.

Title of abstract: The new Istat quality policy for statistical production: challenges for a new data ecosystem

Gabriele Ascari, Maria Francesca Loporcaro, Mauro Scanu, Giorgia Simeoni – Istat, Rome, Italy
Antonia Boggia – formerly at Istat, Rome, Italy

Since the 90s Istat has adopted a systematic approach to ensure the quality of statistical information. In the years 2016–2020 Istat carried out a modernisation process whose main aspects were the centralisation of some phases of the production process, and the set up of the Integrated system of statistical registers (SIR) as a base for the corporate statistical production. In addition, a great investment is being made at Istat on the statistics based on new data sources with particular reference to Trusted Smart Statistics (TSS).

Due to these organisational and strategic changes, the approach to quality had to be revised for the new data ecosystem.

In 2020 Istat top management renovated the Institute quality commitment through the creation of the role of Quality manager and the re-constitution of the Quality Committee that was already active in the years 2010–2016, assigning them the task, among others, to prepare a proposal for a new Istat policy on quality for the statistical production.

Such a proposal has been produced and then approved by Istat top management in October 2021. It includes a set of activities to be carried out in the next 3 years.

The new policy deals with the main aspects of quality in official statistics production, and for each of them illustrates the state-of-the-art and the planned improvement actions.

Obviously, the European Statistical System (ESS) common quality framework is adopted as a reference. Then, the first aspect that is described refers to the tools for quality assessment. Istat relies on quality guidelines for surveys and for statistics based on administrative data; quality reports following ESS standards are also available. Metadata systems documenting the quality of administrative data and statistical processes, enriched with standard quality indicators, are constantly updated, but there is a need of renovating the national metadata system to include both structural and reference metadata and to be able to satisfy the needs of different stakeholders.

The core of the new policy is the proposed procedure for the quality assessment of statistical processes. It is differentiated for type of process. A checklist to verify conformity of traditional statistical processes to sound methodologies will be applied to label compliant processes, while a system of quality indicators is being defined for the processes of the SIR. A quality framework for statistics based on new data sources and TSS should be established.

The new quality policy takes also into account the different stakeholders of statistical production, such as the users, the data providers and the internal staff.

Finally, attention is posed on the coordination and support to the National Statistical System, that includes Other National Authorities (ONAs) producing European Statistics and other bodies. A programme of audit has been carried out on the ONAs while a specific quality assurance framework is under development for the others.

Keywords: quality assessment; quality policy; checklist; labelling.

Title of abstract: Managing quality assurance

Anne Kaag Andersen, Karin Blix – Statistics Denmark, Copenhagen, Denmark

Traditionally quality assurance has been dealt with decentralised in Statistics Denmark. There seems to be consensus of some minimum tasks to ensure quality, but these are not written down in some common guidelines for the organisation. For the past seven years, Statistics Denmark has conducted internal reviews on central statistical products. In these reviews, recommendations on quality assurance procedures have been given, but there has not been a fixed template for quality assurance in place. Statistics Denmark have decided to make use of GSBPM to have a common terminology for statistical processes in Statistics Denmark. This paper discusses what minimum standard there should be for quality assurance for different modes of data collection and what quality assurance tool can be used across domains and collection modes. Examples of check lists for minimum requirements will be given and discussed.

Keywords: Quality assurance, Quality management, Code of Practice, GSBPM, Continuous improvement, Minimum requirements.

Session 12
MACHINE LEARNING AND TECHNOLOGIES FOR EFFICIENCY

Title of abstract: Early provision of economic short-term indicators using Machine Learning

Sandra Barragán, Jorge F. Calatrava, Juan Carlos Gálvez Sáenz de Cueto, D. Salgado – Dept. Methodology and Development of Statistical Production, Statistics Spain (INE), Madrid, Spain

Lasai Barreñada – Dept. Statistics and Operations Research, Complutense University of Madrid, Madrid, Spain

José Manuel Martín del Moral, E. Rosa-Perez – S.G. for Industrial and Services Statistics, Statistics Spain (INE), Madrid, Spain

Short-term business statistics (STS) are the earliest statistics released to show emerging trends in the European economy. Their relevance is reflected in the fact that 7 out of the 27 indicators included in the European Statistical Recovery Dashboard are STS indicators. The major advantage of the monthly and quarterly released STS data is that they are available very shortly after the end of the reference period. Nevertheless, during the Covid-19 pandemic, the National Statistical Institutes (NSIs) were asked to improve the timeliness of these indicators in order to have faster data that could be useful for policy making. Inevitably this must be carried out searching for a trade-off between speed and quality.

Many methods and data sources can be used to obtain advanced indicators that take into consideration accuracy issues. Here we present a process that allows for obtaining the Spanish Industrial Turnover Index (ITI) at the very first moment that data are available, around 20 days after the end of the reference period. This amounts to a notable reduction from the current 51-day release period. We follow the philosophy of the GREG and Sanguiao-Zhang estimators to propose, specifically for a cut-off sampling design, a combined projective-predictive estimator. In this combination collected data are plugged-in in the estimator whereas not yet collected data are predicted.

The predictions are done by using Machine Learning techniques to estimate the yet uncollected values based on regressors built with microdata and paradata from the past and the current reference period of the ITI as well as of other short-term statistical operations closely related to ITI. The algorithm used so far has been the light gradient boosting on regression trees.

The improvement in the timeliness quality dimension is obvious but there is a strong compromise to measure the accuracy as well. We use conventional techniques to estimate the root mean square error, which is computed as the figure of merit to assess how close the advanced estimates are to the final statistics.

Promising preliminary results have been obtained for ITIs for different geographical and economic breakdowns. Estimations of the ITIs on data before, during and after the pandemic have shown positive outcomes.

Some preliminary conclusions from this experimental exercise are:

– Statistical learning algorithms with high predictive capacity on early data of the reference time period allow us to improve timeliness under a controlled compromise of accuracy;

– Both microdata and paradata are relevant for high-quality predicted values;

– Outliers are extremely hard to model and predict and they clearly need management and processing by subject-matter experts;

– Measurements errors have an important role in a good quality estimation so they have to be taken under control;

– High-quality experimental statistics can be produced with traditional survey data by using novel statistical methods in Official Statistics.

Keywords: Machine Learning; Short-Term Statistics; timeliness; accuracy; relevance.

Title of abstract: A Universal Automated Rules-Based and Machine Learning Service for Occupation and Industry Classification of Surveys

Alex Howard, Lucy Furgussonm, Chris Brooker, Michaela Morris, Claire Griffiths, Rosie Maslin, Suzanne Fry – ONS, UK

Matching data, including those collected for statistical and administrative purposes, to the UK Standard Industrial Classification of economic activities (SIC) and UK Standard Occupational Classification (SOC) frameworks is vital for developing key statistics, for comparison with international trends, and for informing UK policy. Within the UK Office for National Statistics (ONS) this has been achieved through a combination of automatic matching using tools such as G-Code, developed by Stats Canada, with the remainder requiring costly manual matching. Implementation of automatic matching is not consistent across the UK. These differences have led to matching results with highly variable quality. The need for a universal, high quality, consistent automated service that can be accessed securely by both the ONS and external stakeholders is self-evident.

The ONS Classifications team manages the SIC and SOC frameworks. We were asked to produce an in-house automated matching solution to sit within the processing pipeline for Census 2021, as G-Code cannot operate within our secure environment. We developed a rules-based tool in Python, and curated reference databases, which were able to match 64% of responses to SOC and 57% to SIC with an accuracy of over 90%. These results were much better than the 47% match rate achieved by G-Code on similar datasets. The tool was highly efficient, able to process the full Census in a matter of hours and run multiple instances in parallel. We also developed monitoring tools that allowed us to adapt the reference databases and tune thresholds to adjust for better quality during live operations to improve performance in real time.

The ONS Data Science Campus carried out a simultaneous investigation into using machine learning to match SIC and SOC codes. Their results indicated a logistics regression algorithm was able to match codes to 77% of entries with a 90% accuracy

against Census data, although these metrics were highly sensitive to the quality of the responses. This method also allowed for a high degree of tuning by the user to adjust quality and was highly versatile.

Due to the wide demand across UK SIC and SOC users for a central automatic matching solution, Classifications began a research project in 2021 to compare and improve both the rules-based and machine learning tools, testing them against different datasets and under different conditions. The aim was to produce a recommendation for a single automated SIC and SOC Matching Service for use across the ONS and UK government. This resulted in a recommendation to develop a hybrid solution, which combines the certainty and stability of matching against reference databases, with the versatility and adaptability of machine learning.

We present the results of this recommendation and introduce our new hybrid tool for use not just by ONS, but across government to improve interconnectivity, quality, and consistency of data processing better suited to the Digital Age.

Keywords: Classification; Automation; Machine Learning; Algorithm Development; Research and Development.

Title of abstract: Robotic Process Automation (RPA) in official statistics: an NSI use case and lessons learned

Eszter Nagy, Csaba Füleki, Ferenc Csanádi, Sándor Tokai, Andrea Petres, Zoltán Vereczkei – Hungarian Central Statistical Office, Budapest, Hungary

Improving efficiency is an ever-present challenge for all organisations around the world, with the National Statistical Institutions (NSIs) being no exceptions. Efficiency in NSIs are interpreted on three levels: product, process level and also on the level of the organisation. In a world putting more and more pressure on NSIs to produce more timely statistics with no real compromises in other quality components, the push to be more efficient is present for all processes of an NSI, statistical business processes and other, non-statistical supporting processes as well.

One way an NSI can be more efficient is to use process automation in order to speed up processes and to free up human resources and reallocate those to other areas where the value added in the terms of quality is higher. Introducing automation is also a mitigating action as it minimises the risk of human errors during the processes. In this sense, automation also has extra added value in quality for an organisation.

For statistical business processes, this automation is usually ensured by different forms of automatic solutions for statistical activities such as validation checks, data cleaning or coding. For non-statistical supporting processes automation is motivated by the similar goal: to reallocate resources into processes that has more added value in terms of quality while increasing the number of exact processes run under the same amount of time with no compromise in quality. Robotic Process Automation (RPA) is a universal solution to ensure that repetitive, well-defined processes with low operation risk can be fully automated and can be managed by robots only, with no human intervention, apart from the management of the robots themselves.

Based on information on current practices gathered from international organisations, the Hungarian Central Statistical Office has started experimenting with RPA solutions. Collecting information on the theoretical background, the technical perspectives was the necessary first step in this work but as part of this pilot, actual processes were also selected, described and transformed into RPA solutions. The journey with both the preparatory and implementation phases of this work was very fruitful to gather first-hand experience with the technique and to become capable to recommend a more generic practice to be followed with the RPA in the NSI in the future. These experiments covered processes that are more of a supporting process nature but processes that are more closely connected to the statistical business processes have also been experimented with.

Based on the experience gathered during these first experiments, the Hungarian Central Statistical Office would like to share the lessons learned and some recommendations for statistical organisations on designing such experiments, the practical experience we have gathered with our pilot so far and moving forward with more integrated RPA solutions in NSIs.

Keywords: RPA, process automation, efficiency.

Title of abstract: Machine Learning Process in the Production of Statistics

Marek Rojíček, Jaroslav Sixta – Czech Statistical Office, Prague, Czech Republic

Modern statistics is facing challenges that were hardly expected a few years ago. Discovering new data sources useful for official statistics called for new methods of data processing that will be fast and efficient. Moreover, the issue of big data counting thousands or more records received by statisticians every day, week, or month is impossible to handle by traditional techniques used in the official statistics. The Czech Statistical Office was confronted with big data for the first time by obtaining data from retail chains. This data contains bar code classification, sales, quantities, and a set of attributes provided by each retail chain. According to the contracts with private data providers, the Czech Statistical Office receives twice a month data sets for the most important retail chains, drugstores, pharmacies. Currently, hobby markets, furniture and clothing stores are planned. The crucial question of the data processing is the transformation of bar codes or an internal classification into a statistical classification, namely COICOP (Classification of Individual Consumption by Purpose) and CPA (Classification of Products by Activity). Without this classification, data processing cannot be done. With respect to the thousands of new products introduced on the market every month it seems impossible to classify these products manually within a few days. The solution was found in the process of machine learning where the translating database is being filled automatically by logistic regression. This process has higher than 90% reliability and only few cases need deeper look of statisticians. Currently, this data is successfully used for three main purposes: monthly consumer price index, quarterly household consumption, and monthly trade statistics.

Keywords: machine learning; consumer price index; logistic regression; big data.

Title of abstract: Monitoring sub-national longevity and mortality disparities: challenges and opportunities

Domantas Jasilionis – Max Planck Institute for Demographic Research, Rostock, Germany, Demographic Research Centre, Vytautas Magnus University, Kaunas, Lithuania
László Németh, Dmitri Jdanov – Max Planck Institute for Demographic Research, Rostock, Germany

Health and mortality disparities are one of the key challenges for sustainable longevity improvements at the national level. Therefore, comprehensive and reliable data beyond national rates are needed for strengthening evidence-based policies aiming at reducing health inequalities. Monitoring mortality changes at subnational levels provides evidence about country-specific development in the context of health transition. There are examples when the stagnation or reversal trends in mortality in deprived areas have been responsible for the stagnation of life expectancy at the national level. Geographical mortality disparities within countries might be much higher than observed mortality disparities between countries. In this study, we are focusing on spatial inequalities in longevity and mortality and their relationships to area-level socio-economic characteristics.

There are substantial data and methodological challenges for measuring sub-national longevity and mortality. Even though, modern statistical systems can produce enough detailed data to study subnational mortality at any level of detail. Nevertheless, there are two principal problems: on the one hand, restricted access to detailed data because of the data protection rules, and on the other hand, stochastic uncertainty because of the small numbers of events. However, there are ways how to make some use of area-level data in the framework of health inequality research such as aggregating areas into the few socio-economic groups according to area-level characteristics (e.g. deprivation index).

The problem of random noise related to small area-level populations and deaths might be solved using an aggregation over age groups and/or time intervals. This traditional demographic approach has several drawbacks. First, aggregation over time leads to less precise information about the temporal changes, e.g. combining the years 2019 and 2020 would lead to imprecise judgement about the pandemic-related mortality. Second, in low mortality countries, aggregation over age may require using broad age groups that are not acceptable for the calculation of life tables or age-standardized indexes. An alternative modern state-of-the-art approach is based on using the hierarchical Bayesian model. Although it provides a reliable picture of the overall spatial mortality variation, some of the details might be still missed due to smoothing. In addition, the aggregation across small areas would not necessarily fit the national mortality pattern.

In our paper, we propose a framework that combines the two approaches to get reliable and consistent estimates of sub-national longevity and mortality at various levels of subnational division in Lithuania.

Keywords: Covid-19; weekly excess mortality, monitoring pandemic.

Title of abstract: Linking the Swedish Labour Force Survey to compensate for time series breaks caused by the new EU framework regulation

Thomas Önskog, Frida Videll – Statistics Sweden, Solna, Sweden

The Swedish Labour Force Survey (LFS) is the foundation for official statistics on the Swedish labour market. It is the only source of statistics that continuously provides a coherent picture of the labour market in terms of employment, unemployment, hours worked, etc. and the interest from users is therefore considerable. As from January 1, 2021, the LFS must comply with the new EU framework regulation on social statistics. To do so, several changes has been made to the survey, both regarding the population and the questionnaire. In addition to the changes caused by the new framework regulation, a revision of the auxiliary information used in the estimation was implemented at the same time as the new framework regulation. To the serious concern of the users, the changes made to the LFS has caused breaks in many of the time series.

To estimate the size of the time series breaks and, by extension, to link the time series of the LFS, Statistics Sweden has performed a parallel run with both the old and the new questionnaires during 2021. Unfortunately, the modest sample size for the old questionnaire in combination with the limited duration of the parallel run leads to considerable uncertainty in the results from the parallel run. This makes it hard to draw conclusions from the parallel run and, consequently, this data must be complemented by data from a series of other sources. The most important of these other data used in the analysis are data on the number of respondents that are affected by a change in the definition of employed (and long-term absent), estimates based on both the old and the new auxiliary information, respectively, as well as flow data describing how the employment status of respondents vary from one quarter to the next.

In this paper, we describe how we have analysed and combined the available data to derive estimates of the size and sample variance of the time series breaks. We also describe how we use these break estimates to construct linked time series extending back to 2005. These linked time series will be disseminated to the users in order to meet their need for time series. The main tool in the analysis is a suitable choice of weighted means. In addition, we use models for seasonal adjustment to verify that no time series breaks remain in the linked time series.

Keywords: labour force survey; time series; linking; user demands.

Title of abstract: Bias-variance trade off on the use of non-response weights in inequality estimates

Josep Espasa Reig – LIS Cross-National Data Center in Luxembourg, Esch-Sur-Alzette, Luxembourg

The Luxembourg Income Study Cross-National Data Center (LIS) is a research center devoted to cross-national analysis of income and wealth data. After the data is acquired, harmonized and documented, LIS produces a series of estimates showing inequality measures for many different wealth and income outcomes (up to 39 years for 52 countries).[3]

The data acquired by LIS typically contains sampling weights computed at household-level. In recent years, some data providers have started producing additional weights at individual-level. These contain an extra adjustment by non-response propensity of individuals. When compared to household-level weights, individual-level weights increase the representation of males, immigrants, one-person households, non-home owners and, in recent years, people aged 24 to 34. The impact of using individual-level weights on income, poverty and inequality indicators is currently unexplored.

In the last decades, falling survey response rates are a widespread trend that has prevailed across developed countries.[4] Lower response rates are associated with higher non-response biases,[5] and thus require larger weighting adjustments.[6] In this context of increasing variance in non-response weights, there is a risk of over-adjusting samples. In such case, benefits from bias reduction could be partially and even totally neutralized by the increasing standard errors.[7] In certain situations, using non-response adjustments could even be counter-productive for the quality of estimates. This could happen if the adjustment is modelled using auxiliary variables not related to the dependent variables in the estimate.

---

[3] Examples of these are Gini and Atkinson indexes, mean and median incomes, poverty rates, etc.

[4] E.g. Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018, April 20). Response rates in the European social survey: Increasing, decreasing, or a matter of fieldwork efforts? Survey Methods: Insights from the Field (SMIF). Retrieved November 30, 2021, from https://surveyinsights.org/?p=9673

[5] Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. Public Opinion Quarterly, 72(2), 167–189. https://doi.org/10.1093/poq/nfn011

[6] Little, R. J. A. (1986). Survey Nonresponse Adjustments for Estimates of Means. International Statistical Review / Revue Internationale de Statistique, 54(2), 139. https://doi.org/10.2307/1403140

[7] Pike, R. G. (2022). Adjusting for Nonresponse in Surveys. In Higher Education: Handbook of Theory and Research (1st ed. 2021 ed., pp. 411–449). Springer. https://doi.org/10.1007/978-1-4020-5666-6_8

We measure the extent to which individual-level weights reduce the estimation bias and increase variance in LIS datasets, as compared to those obtained with household-level weights. To estimate bias, we benchmark our survey estimates for wages and income with those provided by the OECD National Accounts. We then measure variance inflation by computing the Design Factor, Design Effect of the weights, together with the Standard Error of LIS estimates.

The analysis uses 35 LIS datasets from Germany (originating from GSOEP) and 31 from the US (originating from CPS-ASEC). These two countries are selected because they fulfill two criteria: their individual-level weight has a non-response component; and they have time-series long enough to observe how bias and variance have changed across years. The variance estimation is performed using bootstrap for the sum, mean, median, Gini coefficient and Atkinson index. The process takes into account the survey design (i.e. multi-stage sampling and stratification variables) whenever there is detailed enough survey information.[8]

---

[8] Following the techniques in: Wolter, K. (2007). Introduction to Variance Estimation (1st ed.). Springer. https://doi.org/10.1007/978-0-387-35099-8

SESSION 14
METHODOLOGICAL INNOVATION FOR QUALITY IMPROVEMENT

Title of abstract: More data sources, more information, more quality

Sofia Rodrigues, Almiro Moreira, Paulo Saraiva – Statistics Portugal, Lisbon, Portugal

The increasing and intensive use of non-statistical data sources (administrative and other) brings new and important challenges to Statistics Portugal and the official statistics production processes.

Statistics Portugal, in addition to its own infrastructure and extensive experience in data capture, has already created a robust set of technical components: Data Warehouse infrastructure, Survey Management System (SIGINQ), Population and Samples System (SIGUA), the centralization and standardization of collection processes, metadata systems and of statistical methods are some of the examples. However, the system for compiling economic statistics, notably in terms of editing and analysing information, is still strongly supported by a bottom-up approach, with an emphasis on microdata. In addition, data analysis and quality control processes have focused more on data traditionally collected through surveys than on data from external sources.

In the context of Statistics Portugal's strategic objective for the creation of the National Data Infrastructure and in view of the increasing availability of various types of data, statistical and non-statistical, collected or not for statistical purpose, and the consequent appropriation and use in the production of official statistics. It is (even more) urgent to reassess the operating logic of the statistics production process in an integrated manner, from data collection, cleaning, editing, processing and analysis, to dissemination. Recent organizational rearrangements at Statistics Portugal facilitate a change of the production chain towards a data-driven perspective.

The main objective of this paper is, taking as a reference the models developed in other countries (namely the Netherlands and Ireland), to suggest a transition from a bottom up approach to the analysis and processing of the various data types to a top-down logic, in which microdata is verified but based on reference at a more aggregate level: If macro analysis indicates unlikely results, the individual data responsible for them will be verified and eventually edited. Results expected are, on one hand, an increase in efficiency, saving time and resources by reducing the volume of microdata to analyse and / or editing, and on the other hand, an increase in quality, as the results to be published are one of the aspects to be aware of at the outset.

Keywords: non-statistical data sources; top-down data analysis; data-driven production processes.

Title of abstract: Flash estimation of gross value added in Wholesale and retail trade using combined data source

Klaudia Máténé Bella, Ildikó Ritzlné Kazimir, Timea Cseh – Hungarian Central Statistical Office, Budapest, Hungary

The flash estimation of quarterly gross value added of the Wholesale and retail trade in Hungary is based on different data sources, namely short-term statistics data, return trade surveys, online cashier data, VAT data, online invoices data and social contributions data. The quality of the calculation is important because this section had a significant share (9.1%) of the total economy in Hungary in 2020. We argue that the quality in this case can be measured by the revision after the benchmarking of quarterly data to annul data. The different statistical and administrative data sources have different coverages and include different measures. The main problem is that the direct estimation of gross value added of Wholesale and retail trade is not feasible based on these data separately. The estimation can be built only on a combined data source.

In order to solve this problem, we analyzed all data sources. First of all, we mapped their relationships. The coverage of these data sources is varied in the time of flash estimation. Monthly and quarterly VAT returns and statistical surveys are submitted by large companies, but all companies having employees have to submit social contributions returns, too. The records of these databases are tax units, not statistical units, therefore we had to use data of Business Register to determine the main activity according to NACE Rev. 2. Using the VAT data, only turnover, not gross value added can be calculated directly for companies. The national accounts' concept should contain the trade margin. Therefore, we calculated the difference of payable and deductible VAT base and after outlier detection, we analyzed the relationship of this difference and the production value. We found that this relationship can be forecasted by econometric method and it improves the quality of estimation because the development of inventory is taken into consideration also for smaller companies. This companies usually don't have inventory data or if they have than data are very unreliable.

We made our econometric estimation using times series based on combined data for the year 2020. It was a quasi-real time estimation and we concluded that the revision between the quarterly data and annual data can be reduced using this model.

In future, we are planning to extend our combined data sources with data of online invoices system which has a full coverage because every company (including sole proprietors) is subject to this system.

Keywords: flash estimation, combined data, multi-source statistical system, wholesale and retail trade, national accounts.

Title of abstract: Machine learning based quality assurance in web scraping of exchange rates

Andreas Dietrich, Andreas Rehs – Deutsche Bundesbank, Frankfurt, Germany

The Deutsche Bundesbank currently collects and disseminates official bilateral exchange rates of the Euro and the US-Dollar vis-à-vis 195 currencies. While for about 40 countries exchange rates are gathered on a daily basis, the remaining currencies are available on a monthly basis (Deutsche Bundesbank, 2020). This high-quality database partially reaches back to 1940 and is used by a number of European partners such as Eurostat or various national central banks within the ESCB, by national federal ministries, by judicial authorities, by public administrations such as customs and statistical offices as well as by numerous enterprises and private users. Currently, the Deutsche Bundesbank collects exchange rates, to some extent manually, from monetary authorities via electronic media (internet or corresponding sites of economic information services). We intend to investigate to which extend automatisation tools such as web scraping can assist the data collection process or even replace parts of it. For the quality assurance and the improvement of the web scraping process itself, we are furthermore going to implement a machine learning approach. This shall be used for plausibility checks and for the identification of errors in newly web-scraped data.

In our setup, we are going to choose a selection of the 195 websites of foreign central banks and scrape those for daily exchange rates for a period of two months beginning in January 2022. At the same time, we aim to control whether the scraped values align with those collected from a human coder. We label correct scraping and scraping errors accordingly and classify the latter to different categories. In a next step, we are going to use this dataset in combination with the currency-specific time series to train a set of machine learning algorithms like random forest, logistic regression and a neural network for the identification of errors. Afterwards, the algorithms shall be tested on newly scraped and manually controlled values.

Our results shall reveal whether the web scraping-based collection of exchange rates provides a viable alternative to the current collection process. Furthermore, we want to gather knowledge on the sources of errors in web scraping. Additionally, we are going to examine whether they can be reliably identified with machine learning algorithms.

Keywords: exchange rates; web scraping; machine learning; quality assurance.

Literature:

Deutsche Bundesbank, 2020. Quality report on exchange rate statistics.
Retrieved 10.12.2021 from:
https://www.bundesbank.de/resource/blob/813896/7255d06454f3445376b980c360c2d414/mL/qualitaetsbericht-data.pdf

Session 15
METADATA INNOVATIONS

Title of abstract: The value of the French metadata repository: a producer's perspective

Xavier Helfenstein – Insee, Limoges, France

The Permanent database of facilities (BPE) precisely identifies and geolocates different equipment, services and infrastructures which are accessible to the general public throughout France. This database is produced by collecting many administrative data sources. Each year, at the end of the production process of a vintage, various data files are provided to different recipients internally at INSEE or to external partners.

The diversity of the delivered products and the large number of their recipients require the production and updating of a substantial documentation, which needs to be available in various formats. As a result, a wide range of metadata is associated with each product: dictionary of variables, file drawing, release note, explanatory note, classification of facilities types, etc. This metadata differs in both content and format for different recipients. Taken together, they constitute a heterogeneous and voluminous set whose updating with each vintage of the database requires an excessive workload if it is not rationalised.

A BPE-specific metadata system was therefore designed to optimise these updates. It was built using tools developed as part of the implementation of the statistical metadata repository (RMéS) at INSEE. Abundant statistical metadata is thus structured according to a standard adopted by Eurostat: the SIMS format (Single integrated metadata structure). This information is supplemented by structural metadata, relating to the disseminated variables and their related modalities, designed to meet the documentation needs of their recipients, or to be used as inputs to their processes. For example, the release notes or the additional explanatory notes are made up of RMéS metadata elements.

At each vintage of the database, the producer makes use of the metadata system to design and provide the needed information to each recipient. This system is really valuable in terms of:

– coherence → the products and data are being consistently described using the same repository;
– cost effectiveness → the updating load is significantly reduced, being made once for all purposes;
– convenient → updating is made in more ergonomic interfaces that facilitate the structuring of the information.

Keywords: metadata; system; facilities; database.

Title of abstract: STATMIK – A multipurpose application that makes management of reference and quality related metadata a pleasant job

Rudi Seljak, Tina Steenvoorden – Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia

The Statistical Office of the Republic of Slovenia (SURS) has a long tradition of producing and using standardized reference and quality related metadata. In the past decades a lot of this information has been collected, stored and to a certain extent also disseminated, making it accessible to various (internal and external) users. As the amount of these data increased over time, the problem of their control and the possibility of comprehensive analysis became more and more prominent. The information was scattered in different locations and in different forms, which significantly reduced its analytical power.

To overcome this problem, we started to develop a new, multipurpose application that would enable easier and more effective usage of reference metadata produced through the statistical process and would support the evaluation phase of our statistical business model. The application was developed for several years and came to its production life in 2019. The main functionalities of the application are:

- All the collected reference and quality related metadata are stored in one, central database and are ready to be used for different purposes.
- Most of the quality indicators are calculated by automated procedures, which store them in the central database.
- The application enables automated creation of Methodological Explanations and Quality Reports based on the information stored in the central database.
- Comparative analyses of quality indicators, including comparisons through time and across domains, are enabled by the application.

The paper describes the integrated architecture of the application, some further details of its functionalities, and points out main challenges that we met during its development. We also provide experiences from the first years of its usage and sketch some ideas for further development.

Keywords: reference metadata, quality indicators, quality reporting, metadata, standardization.

Title of abstract: A step towards metadata-driven production

Rohan Draper, Ole Schnor, Charlotte Nielsen – Statistics Denmark, Copenhagen, Denmark

Standardisation is key to enable metadata-driven production in official statistics. This paper describes how quality assured standard classifications from one system, can be used to data storage and data validation, editing and production of tailor-made analysis.

To be truly metadata-driven, by feeding metadata into the statistical production, production processes must be defined in a unified way. For this, we use the Generic Statistical Business Process Model (GSBPM).

Metadata – in this case classifications – are stored in a metadata repository as information objects, in accordance with the Generic Statistical Information Model (GSIM). From this metadata repository, quality assured standard classifications are used in the GSBPM sub-processes 5.3 Review and validate and 5.4 Edit and impute. The classifications are also used to produce publications and tailor-made analyses.

Keywords: Metadata, Metadata-driven, Standardisation, Classifications.

Title of abstract: Metadata driven data collection: the five next years

Romain Tailhurat – Insee, Toulouse, France

Insee launched a few years ago an ambitious program: to design a new data collection system with metadata at its core. Several milestones have been passed during this timespan: describing questionnaires using a standard representation (DDI), generating business survey instruments from this metadata standard (using Eno, the home-made generator), providing to survey designers a UI abstracting away DDI and allowing direct generation of survey instruments. A lot has been done but more is on the way: the data collection system is progressively supporting household surveys needs, with interviewer platform and mixed-mode capabilities being deployed. Furthermore, a new standard is added to the mix: the Validation and Transformation Language (VTL), as a mean to express dynamic behaviour and data validation. Those new developments are keys to the unfolding of the initial vision: building an integrated and consistent metadata driven collection system.

Keywords: metadata; data collection; survey.

Session 16
VALIDATION AND TECHNICAL ADVANCES, IMPROVEMENT

Title of abstract: Quality assurance from an internationally standardized and generic data validation ecosystem

Olav ten Bosch, Mark van der Loo – Statistics Netherlands, The Hague, The Netherlands

One of the challenges in official statistics is monitoring data quality. Data characteristics observed when designing a statistical process may not continue to hold during production in the long run. If these data flaws are not detected in time, this may lead to costly recalculations or – even worse – undetected errors. It is even more of a problem with new indicators that combine traditional data (surveys, registers) with other, more volatile, new data sources such as transaction data, web data or spatial data. Therefore, data has to be *validated* before being used in production. This holds for statistical processes within statistical organisations as well as for cross-organisational processes, such as data reporting from National Statistical Offices (NSIs) to international organisations.

With the growing amount of official statistics being produced in the European Statistical System (ESS) and the increased demands on timeliness, there is no other way then to *standardise* and *automate* data validation where possible. There have been multiple international projects working on this, resulting in a handbook on validation [1], a set of data validation principles [2], a standard validation report [3] and a list of the most common types of validation rules in the ESS [4]. Combined with standardised metadata in SDMX registries, such as the global registry [5] and the Eurostat SDMX registry [6], the main ingredients for a generic data validation ecosystem are present.

NSIs increasingly use the R programming language in their statistical processes. The R-package *validate* is a popular tool to validate data [7]. It offers a language for defining validation rules supporting any type of data checks, including uni- and multivariate checks, statistical checks, checks on time series, hierarchical aggregation, and more. The software can execute analyses on possibly large datasets, providing statisticians with feedback on the health of their data. Validation results are presented graphically and in a machine-readable report, which can be used as input for consecutive processes. Common rules can be re-used among multiple validation processes, which makes it a useful tool to also implement the ESS main types of rules.

In the *ValidatFOSS* [9] project all ingredients were combined to provide a *free and open source generic validation ecosystem* that maximally aligns with ESS standards such as the handbook, SDMX and the main types of rules. The architecture is generic and can be used in any statistical domain, on any SDMX compliant registry and for data

---

[9] ValidatFOSS: Validation with Free and Open Source Software. ESTAT (GAs) No: 825659 and 882817

validation within organizations as well as for international data reporting. An online validation cookbook [8] offers recipes for implementing the most common validation scenarios found in official statistics. In this paper we highlight the results, present some example cases and philosophize about the role this building block can play to monitor data quality in the new emerging data ecosystem of official statistics.

Keywords: data quality, data validation, data reporting, standardisation, R, SDMX.

References

[1] *Methodology for data validation* 2.0, Revised edition 2018, ESS Standard, https://ec.europa.eu/eurostat/ramon/statmanuals/files/methodology_for_data_validation_v2_0_rev2018.pdf

[2] *Principles for data validation*, TF on validation, https://ec.europa.eu/eurostat/cros/content/principles_en

[3] *Design of a generic machine-readable validation report structure*, M. van der Loo, O. ten Bosch, Version 1.0.0
August 15, 2017, https://ec.europa.eu/eurostat/cros/content/validation-report-structure_en

[4] *Main types of validation rules for ESS data*, European Commission – Eurostat/B1, Working Document, ESS data validation project, 2 February 2018, https://ec.europa.eu/eurostat/cros/content/02c-main-types-data-validation-rules-and-fictive-domain_en

[5] *SDMX Global Registry*: https://registry.sdmx.org

[6] *Euro SDMX registry*: https://webgate.ec.europa.eu/sdmxregistry

[7] *Data validation infrastructure for R* , Loo, MPJ van der and E de Jonge (2021), Journal of Statistical Software 97 1-22.

[8] *The data validation cookbook*, Loo, MPJ van der and ten Bosch, O (2020) https://data-cleaning.github.io/validate

Title of abstract: Validation VTL, SQL, R: How One Size fits All

Pedro Cunha, Sónia Quaresma – INE, Lisbon, Portugal

All the data in our NSIs (National Statistical Institutes) needs to be validated. Furthermore, data exchanged with Eurostat must meet criteria, agreed upon previously by the domain groups. The exercise of validating the data is humongous and tends to increase with the addition of new data sources. On one hand the ESS (European Statistical System) members share and agree on the same validation rules, on the other hand they perform their validation on their own. Although Eurostat provides tools, for several reasons, including the existence of national validation rules that use the national validation language and the requirement that the NSIs' data doesn't leave premises this is not always a feasible solution.

Doing the validation exercise on their own means NSIs are isolated in the quality assessment of their validation rules. However, the rules used belong in most cases (ranging from 80 to 95% depending on domain) to a small set of different types that have been identified and gathered in the MAin Type of VAlidation RUles Template (MATVARUT).

Using MATVARUT has enabled us to build generic rules that may be parameterized by the domain expert users. These rules are domain agnostic, and, in this sense, they are used, therefore tested by a larger group of users within different domains. The rules based on the MATVARUT can be shared among ESS members thus enlarging the community that revises and assesses it. For this purpose, MATVARUT based SQL (INE-PT) and R (CBS-NL) rules were placed and maintained in Github, and lately (February 2021) Eurostat has been preparing their own set of MATVARUT rules in VTL.

At the Portuguese NSI the first agnostic domain template-based validation rules were developed using SQL in 2018 for the ANIMAL domain and covered all the rules it used resorting to 10 of the 12 different types of rules envisioned in the template. Some of the rules did not apply to ANIMAL as they concerned time series and seasonal adjustments. This fact made clear the need to apply, test and extend its use to another domain. In 2020 we started a project of validation using the main type of validation rules template in NAMAIN (National Accounts).

The paper explains, using examples, the need to extend MATVARUT both in terms of new types of rules to be added and of subtypes of previously existing rules. The requirements were identified not only in the NAMAIN area but also in OJA (Online Job Advertisements) which webscrapes its data.

Although we are convinced of the need of the enlargement of MATVARUT this needs to be shared and assessed within the ESS so that MATVARUT may on its way to become a standard ensure that despite our use of SQL, VTL, R or any other validation language one size still fits us all.

Keywords: Validation; SQL rules; Quality Assessment.

Title of abstract: Improving statistical data editing with Machine Learning: first use cases)

Sandra Barragán, David Salgado – Dept. Methodology and Development of Statistical Production, Statistics Spain, Spain

Statistical data editing is a crucial step in the end-to-end statistical production process in any National Statistical Institute (NSI) to warrant data quality (especially accuracy) in the final outputs. However, this production phase entails resources, both human, financial, and, more critically, time. The Generic Statistical Data Editing Model (GSDEM) provides a comprehensive framework comprising multiple business functions (review, selection, treatment) to detect and correct multiple non-sampling errors. Statistical data editing is also critical from the quality assurance perspective since it impinges on cost- efficiency involving thus the adequacy of resources.

There exist different data editing modalities dealing with diverse aspects of this process phase, namely, domain editing, editing systematic errors, selective editing, interactive editing, automatic editing, and macro editing. Selective and macro editing modalities are clearly oriented towards cost-efficiency. We have focused on these process steps using Machine Learning techniques.

The use of Machine Learning techniques for the editing phase of the traditional statistical production process has received notable attention in the last years (UNECE, Workshop on Statistical Data Editing, 2020). Due to its automatic character in nature, these techniques meet perfectly the idea of using innovative and appropriate statistical procedures if they are applied in an adequate manner with a transparent implementation as part of a modular and standardised process.

We use Machine Learning techniques for the following aspects:

– To compute item (local) scores to rank variables with influential errors in both continuous and categorical variables, thus bringing editing workflows of household and social statistics closer to those of business statistics. This simplifies the generic editing framework. This is accomplished using random forest models.

– To process also semicontinuous variables, both for error detection and error treatment (imputation), either in business statistics and in social statistics. This is accomplished with a two-step combination of a classification model and a regression model using random forests.

– To deal with imbalanced data (arising from the low proportion of measurement errors) by analysing oversampling, subsampling and cost-sensitive learning approaches.

– To provide a quality indicator for the unit ranking in both selective and macro editing modalities using a similar approach to the Area Under the Curve (AUC) metrics in Machine Learning.

– To use natural language processing (NLP) techniques to process collection paradata and to construct regressors for selective editing models.

– To use random forests and boosting algorithms to impute yet uncollected continuous variables to produce advanced short-term business statistics.

These are first steps in a systematic study of new methods in the whole statistical data editing phase. This innovation impinges strongly on the quality dimension of cost-effectiveness and the process quality as a whole.

Keywords: Machine Learning; data editing; cost-effectiveness; process quality; selective editing.

Title of abstract: Eurostat's quality assessment of balance of payments statistics in the EU

Leisch Robert – Eurostat, Luxembourg, Luxembourg

The growing importance of statistical evidence, data and information is reflected in the term "evidence-based decision making". Entrepreneurs, politicians and researchers should make decisions based on provided high quality data. The size of bilateral asymmetries in international trade in goods or services is one fundamental indicator related to the quality of data on economic relationships between partners.

Large persistent bilateral asymmetries in international trade undermine the credibility in the provided data by users, could lead to scrutiny of crucial economic indicators like GDP, GNI and may even impact the assessment whether a county is considered to be a net exporter or net importer of goods or services via another country or even the rest of the world.

Ongoing work to better understand bilateral asymmetries, and to reduce them if possible, is typically a task to be coordinated by international organisations.

Eurostat analyses bilateral asymmetries to assess data consistency and if their compilation is made according to internationally agreed reference standards. The current focus is on analysis of Intra-EU asymmetries to improve methodological consistency among EU compilers.

Eurostat's asymmetry management tool builds upon regular qualitative outputs (e.g. Eurostat's quality reports and related recommendations) and quantitative indicators from the data production cycle. Those indicators illustrate the size of the recurrent and largest bilateral asymmetries. The management tool facilitates the preparation of actions together with compilers to reduce asymmetries.

Bilateral asymmetries are at the center of any qualitative action when it comes to comparing international transactions and their origins deserve to be well understood.

This paper describes the methodological and conceptual approach to asymmetries that has been developed in close cooperation with the Eurostat Balance of Payments Working Group, and highlights key results. Further, it proposes further steps to be taken by Eurostat, in close cooperation with countries.

Keywords: Eurostat; Balance of payments; Asymmetries

Session 17
STANDARDS AND STANDARDISATION

Title of abstract: Towards Open and Linked Data Statistical Classifications

Christine Laaboudi-Spoiden – Eurostat, European Commission, Luxembourg

Since the beginning of the millennium, Ramon, Eurostat's Metadata Server has been playing the role of server for the Reference and Management of the nomenclatures and classifications maintained by Eurostat. The classifications and their additional files (correspondence tables, translations, list of official classifications decisions) are available for download under an open license (CC BY 4.0) in various open or proprietary formats and not harmonised in their structure.

This presentation will provide a framework for the standardisation and the conversion of Eurostat multilingual statistical classifications and their correspondences from Ramon, 1) as Artefacts in the Euro SDMX Registry (the Registry) and 2) as Open and Linked Data, with the aim of maximising data interoperability, discoverability and reusability.

In a first step, we developed a conversion tool that transformed the Ramon classifications into Artefacts in the Registry. We based our specifications on the '*Recommendations on Using SDMX annotations*' that define specific annotation-types for formalising the structure and content of code lists and statistical classifications. The Registry will thus store the legacy and new collections of statistical classifications in order to maximise their sharing and reuse across the SDMX Community.

As a second approach, we considered XKOS, an ontology for modeling statistical classifications developed by DDI (Data Documentation Initiative). Although there are similarities with the SDMX annotations related to classification, XKOS is compliant with the semantic web technologies and allows a richer description of the resources as Linked Open Data (LOD). For the purpose of LOD maintenance, we used Vocbench, a corporate semantic tool offered by the Publications Office of the EU for the maintenance of the EU's controlled vocabularies. Each Eurostat classification has been defined in the domain data.europa.eu and a persistent URI (Unique Resource Identifier) assigned for each resource (classification, classification level, classification item, correspondence and association). Based on the persistent URI, we generated correspondences between our classifications, and in addition, to targeted classifications (ISIC, CPC) available remotely in the Caliper platform, a project run by the Food and Agriculture Organization of the UN (FAO). Finally, we linked the classification items to their related classification implementation decisions published in the Official Journal of the European Union, via the European Legislation Identifier (ELI).

A Web access to the statistical classifications is offered via the EU Vocabularies platform or via ShowVoc, the Vocbench human-readable interface. As the statistical classifications are stored in a knowledge graph in RDF (Resource Description Framework), a SPARQL endpoint enables to query the classifications and extract data on demand based on reusable templates.

Keywords: statistical classification; Linked Open Data; interoperability; SDMX annotations; XKOS.

Title of abstract: Data Browser and MetaData Manager: the new dissemination platform for Istat aggregate data

Carlo Boselli – Istat, Rome, Italy

This paper aims to outline the advantages of using a new dissemination tool, the so-called SDMX Istat Framework, a new dissemination platform entirely developed by Istat. The SDMX Istat Toolkit is a set of pick-and-choose building blocks allowing a statistical office to standardize and industrialize the dissemination/reporting business process.

It can be used successfully for building dissemination/reporting SDMX compliant databases and for implementing SDMX in terms of metadata management, machine-actionable dissemination and Open (Statistical) Data. It is composed by several modules (Metadata Repository/Registry, SDMX Web Service, SDMX Metamanager, SDMX Datamanager, SDMX Data Browser) and extends the Eurostat SDMX-RI functionalities, allowing to handle the whole dissemination process in a "metadata and SDMX driven" manner: creation of a new SDMX compliant dissemination database and/or mapping of an already existing database, metadata management, data modelling, data browsing and visualization. The SDMX Istat Toolkit is an open project published on GitHub, and contributions from programmers are welcome. The information search system allows interactive navigation of tables, graphics and dashboards, through a horizontal navigation (based on thematic buttons) and a vertical thematic tree. The text search system returns a list of data flows (queries) which can be further filtered with a subject tree. Dashboards allow simultaneously view of interactive tables and graphs; explanatory text elements, with a renewed and complete set of interactive maps and graphics.

In 2020, the migration of the current dissemination system (I.Stat) to a platform metadata-driven, SDMX-based, began. SDMX is an internationally recognized transfer standard that allows you to easily reuse all I.Stat data and metadata and overcome the technological dependence on OECD, without changing the production and data modeling (CSV) processes. The Permanent Population Census dissemination, which took place for the first time on the new platform, highlighted the potential to manage large amounts of data and metadata in a native SDMX environment ensuring conceptual and functional coherence. The experimentation conducted in 2021, to achieve a complete migration of the Institute's macrodata in SDMX format, highlighted the full compatibility with the statistics production processes and the total reconversion of all metadata, both at structural (DSD, Data Structure Definition) and referential level. Furthermore, the thematic contents will benefit from the new navigation features and an open format display, also available for machine-to-machine web services. The goal is actually to migrate the aggregate data present on the various diffusion systems to this platform starting from I.Stat data warehouse.

Keywords: dissemination system; SDMX; standardization and modernization of the dissemination.

Title of abstract: Enhancing accessibility and clarity by tidying datasets and standardising code lists in the Community Innovation Survey

Flavia Camilleri, Aiva Grinspone – Eurostat, European Commission, Luxembourg
Liliana Avram[10] – Sogeti Luxembourg, Luxembourg

Effective data presentation is a key point in statistical data dissemination: it impacts data readability and usability. In the data management world, the concept of tidy data is used: this allows the proper mapping of the meaning of a dataset to its structure. A principal requirement for tidy data is that each statistical variable forms a column. Datasets should thus be structured so that each variable is a separate dimension, in other words, each dimension contains a single concept. This is not always the case for statistical survey data, where questions can contain various concepts and a thorough study of the data structure has to be performed. Eurostat has launched a project to ensure that the disseminated datasets are properly structured in line with the tidy data concept. To further guarantee that data are easy to visualise and manipulate, one must ensure the quality of structural metadata: the main goal is to harmonise the way dataset variables and levels are coded and to create codes that are meaningful. The standard code lists project, another well-advanced project in Eurostat, ensures that the same concepts are defined in the same way across all statistical domains.

The use of a tidy structure together with harmonised structural metadata brings several advantages for the users: firstly, data presentation is improved and data browsing is facilitated allowing the users to precisely select statistical variables that match their needs. In this way, the users can more easily extract and manipulate the data. Data treatment and processing is also rendered easier in this scenario as statistical variables are conceptually separated. Furthermore, the harmonised structural metadata allow reuse and comparison of data across datasets and domains.

Tidy datasets and standard code lists were successfully implemented in the 2016 Community Innovation Survey (CIS 2016), and the same process has been reapplied in the subsequent CISs. This restructuration project focused on harmonising dataset structures and codes, involved several complex steps and has been made possible by the effective collaboration of several teams of Eurostat. The new way of disseminating CIS data greatly improved their presentation, reduced risks of data misinterpretation, and allowed better re-use of the data.

The project also brought lessons and best practices that can be taken on board for similar projects. The use of tidy dataset structures should be a standard practice for presenting statistical data. Harmonised structural metadata can be ensured through the use of standard code lists. Experience has shown that the highest benefits are reached when these two aspects are combined. This paper outlines the concepts,

---

[10] Working for Eurostat under a service contract between Sogeti and the European Commission (Eurostat)

work performed, improvements made, and next steps for further enhancing the presentation of European statistics.

Keywords: data presentation, standardisation, tidy data, Community Innovation Survey.

Title of abstract: The new updating process of the Italian version of the Nace classification to satisfy users' needs

Francesca Alonzi, Caterina Viviano – Istat, Rome, Italy

Statistical products aim as their main purpose to satisfy user needs and often users complain about their quality above all with respect to the ability of such information to represent timely changes in the reality. The production of high-quality statistics on economic activities carried out by enterprises need an official classification of economic activities harmonized at international level in order to assure coherence and comparability in the statistical system. As it is known, the obsolescence of a classification is determined by the evolving of economies, the faster they move the more often a classification must change.

Istat decided to take action at the national level in order to anticipate the changes later adopted by the multi-year revisions of the internationally driven classification and to update the corresponding national Ateco classification more frequently. To reach this purpose a new organizational model and a new process were set up under the governance of a so-called Ateco Committee coordinated by Istat and composed by national stakeholders of the classification.

The new updating process includes some decision nodes according to a step-by-step approach, starting from the collection and management of proposals for changes to the classification and activation of specific requests, the pre-evaluation carried out by a group of experts of the Ateco classification, the analysis and evaluation by the Ateco Committee and the final approval by the top management of Istat and then by the European Commission. Since the adoption of the Ateco 2007 classification on January 1, 2008 and to date, numerous requests for revisions have been received by Istat over the years, both in the form of changes to existing codes (e.g., greater detail or merging) and in the form of assigning new codes to specific economic sectors of interest. Thanks to the set-up of the new updating process Istat has disseminated the first updating of the classification in January 2021 and it is going to disseminate a second updating next January 2022.

This paper will present the new updating process and the new organizational model.

Keywords: economic activities; Nace Rev. 2; statistical classification.

Session 18
IMPACT OF MEASURING GLOBALIZATION ON THE QUALITY OF OFFICIAL STATISTICS, MNE

Title of abstract: Nationality vs Residency Approach: Measuring the impact of MNEs production structure on corporate financial statements statistics

Walter Susanne, Gerstner Klaus – Deutsche Bundesbank, Frankfurt, Germany

The organisation of global value chains of multinational enterprise groups (MNEs) poses a challenge to official statistics and the underlying concepts of presenting economic reality. At least since the „Irish Case" the awareness about evaluating and addressing these globalisation effects in generating official statistics has increased. Complex, boundary crossing economic transactions and complex ownership structures contravene the concept of measuring national economic production.

We try to evaluate the impact and importance of globalisation on national statistics capturing economic activity of non-financial companies. Therefore, this study uses a combination of information from individual and consolidated corporate financial statements and data about global group structures.

In doing so, we can estimate the part of production that is attributed to the country of interest by applying traditional statistical concepts (residency approach) and compare it with the overall global production of MNEs (nationality approach). This stocktaking exercise helps to understand and shed light on the economic activity that is unobserved by statistics applying traditional approaches.

By exploiting a newly established dataset that combines financial indicators (especially employment, value added, turnover, total assets) and information on national and international subsidiaries of German MNEs, we find that the biggest share of economic production happens abroad. We also benchmark the results with analyses from other countries participating in the stocktaking and other official statistics. Unsurprisingly, the respective global embeddedness of a national economy determines the size of the effect of applying an alternative approach (nationality concept).

Keywords: globalisation; official statistics; corporate financial statements; nationality concept.

Title of abstract: Quality improvements of the European business and macroeconomic statistics: statistical business register as unique and main data source

Agne Bikauskaite – Eurostat, Unit G1

The fragmented picture that statistical authorities of EU and EFTA countries have of multinational enterprise (MNE) groups operating on the European market causes growing harmonisation and quality problems for several types of statistics related to globalisation. With the EuroGroups Register (EGR) project, Eurostat is currently making available a network of statistical business registers used for business and macro-economic statistics production and official statistics quality improvements in relation to coverage on complete set of information, harmonised methodologies and concepts, reduction of burden, cost effectiveness, ensuring accuracy, accessibility, coherence and comparability.

To create the EuroGroups Register, Eurostat receives microdata input on MNE group members (enterprises and legal units) and on their relationships from the national statistical business registers of EU and EFTA countries. The necessary data exchange between the national statistical business registers and European Commission (Eurostat) is defined in the *Regulation (EU) 2019/2152 of the European Parliament and of the Council of 27 November 2019 on European business statistics*. To compensate with worldwide microdata on largest MNE groups Eurostat uses supplementary commercial data source.

This article focuses on the European central statistical business register's quality assessment, quality indicators, quality and process improvements developed and achieved over the time, as well as describes further plans aiming to reach the high quality of the EGR output. In addition, article briefly explains the use of the EGR microdata for production, quality assessment and quality improvement of the official European statistics (such as FATS, FDI, GNI, etc.).

The DGINS emphasised the role of high quality and up-to-date statistical business registers (national and European) as a necessary infrastructure for the coordination of statistical activities and exchange of data on MNE groups. The EGR is part of the European statistical infrastructure and has been built up to better capture globalisation effects as well as for improving the consistency of national data on enterprise groups. The EGR provides the statistical authorities of the EU and EFTA countries with yearly population frames of MNE groups to be used for statistical purpose only.

Keywords: statistical business register, EuroGroups Register, quality assessment, quality indicators, quality improvements.

Title of abstract: Examine of administrative data of Multinational Enterprises for improving the quality of national business statistics

Zsolt Kovari – Hungarian Central Statistical Office, Budapest, Hungary

Since the activities of Multinational Enterprises (MNEs) has increasingly more and more significant effect on the business statistics, several NSIs formed a new unit – called Large Cases Unit (LCU) – in order to examine these MNEs by using various data sources. Although the LCUs follows different operational models and they position in the organisation can be differing, lot of experts agree that LCUs are suitable for ameliorating the quality of business statistics. In the end of 2017 an LCU was also established at the Hungarian Central Statistical Office (HCSO), which was dedicated to handling of MNEs' data and dealing with the relevant administrative data sources.

At this moment the LCU handles VAT data, data of annual reports of the enterprises, data of Country by Country Report (CBCR), online invoice data, as well as partially oversees the receipt of the corporate tax data, but of course it also accesses to the most important business and trade statistics, such as STS, SBS, PRODCOM or FATS.

Online invoice was introduced from the 1st June 2018 and after several correction in compliance with the latest rules practically all of transactions have a record in the online invoice database. HCSO has received the online invoice data from the National Tax Authority by month since January 2021.

These databases allow a comprehensive consistency work, but success of the elimination of the inconsistency depends on the quality of the administrative sources. Therefore, the first step was to examine the administrative data extensively, mapping the relationships among the variables as well as analysing the relations among the statistical and administrative variables. This presentation focuses the VAT data included the data of tax report and the data of online invoice. The LCU had to bring forward the date of data collection which led to a deterioration in quality mainly because of missing data but made the widest use possible. Hence an appropriate imputation process had to be worked out which based on the k-nearest neighbour algorithm.

Beyond this problem the presentation will

– show how has the quality of the VAT dataset improved by the preliminary processing of the VAT data and how can it use for the consistency checks,

– demonstrate the mapping of the business connections using the online invoice data.

Keywords: administrative data; LCU; VAT data; consistency check.

Title of abstract: Addressing large bilateral asymmetries in foreign direct investment flows statistics

Carmen Picón Aguilar – European Central Bank, Frankfurt, Germany
Irene Madsen – Eurostat, Luxembourg, Luxembourg
Morten-Bo Paulsen – European Central Bank / Danmarks Nationalbank, Copenhagen, Denmark

The quality of balance of payments (b.o.p.) and international investment position (i.i.p.) statistics is challenged by the occurrence of differences in the bilateral flows and positions between countries. The compilation of foreign direct investment (FDI) statistics is particularly constrained by the occurrence of complex ownership chains often encompassing multiple entities spanning various countries. The differences in the reporting of FDI transactions and positions between counterpart countries have largely increased in the last years and may hamper the interpretation of the FDI figures.

This paper describes the initiative activated by the Task Force FDI to address in an effective manner the largest outstanding bilateral asymmetries in FDI flows between EU countries. This initiative is known as the Asymmetry Resolution Meeting (ARM), in which national compilers are engaged in addressing the largest intra-euro area/EU asymmetries in a collective manner. Micro data information shared in the FDI network is the basis for the discussion that is enriched with additional metadata and investigations once a particular FDI event is in focus.

The ARM has evolved to become an integral part of the quarterly production cycle of the euro area and EU aggregates prepared respectively by the European Central Bank and Eurostat. The results achieved and presented in this paper show a successful mechanism in prompting national compilers to reduce significantly the largest asymmetries resulting in improved quality of both national and aggregate b.o.p./i.i.p statistics at the EU level.

Keywords: bilateral asymmetries; balance of payments; foreign direct investment; FDI Network

Title of abstract: Web, Paper or Mix? The Impact of Single- and Mixed-Mode Designs on Establishment Survey Participation, Nonresponse Bias, and Costs

Benjamin Küfner, Joseph Sakshaug, Stefan Zins – Institute for Employment Research (IAB), Nuremberg, Germany

The IAB Job Vacancy Survey (IAB-JVS) is a voluntary nationally-representative establishment survey that quantifies the size and structure of job vacancies and other worker flows in Germany. Since 2011, it has been carried out using a concurrent mixed-mode design, with establishments receiving paper questionnaires and the option of online completion. However, this mode design is facing increasing costs and declining response rates. To counteract these trends, a more pronounced push-to-web strategy offers a promising alternative. But the effects of implementing different self-administered survey modes in establishment surveys are less clear than for household surveys. This raises the question, which self-administered mode design works best in terms of response rates, nonresponse bias, and costs in an establishment survey. To test an implementation of an alternative mode design, a large-scale experiment comparing four self-administered mode designs was conducted with 155,000 establishments in the 4th quarter of the 2020 IAB-JVS: Just Online, Just Paper, Sequential Mixed Mode with online invitation followed by paper questionnaire for nonrespondents, and the standard concurrent mixed-mode design.

Further, we experimented with a pre-due-date reminder as an additional response enhancement measure motivating establishments to respond earlier. Response rates, nonresponse bias, and costs per respondent are computed for each experimental group separately. The assessment of nonresponse bias is based on an administrative database covering rich information about establishment and employee characteristics.

The preliminary results indicate that response rates do not differ substantially between concurrent mixed-modes, sequential mixed-modes, and single web modes. The single-mode paper & pencil group has a substantially lower response rate and higher risk of nonresponse bias than the aforementioned three groups. The impact of establishment size on response rates does not differ between the different mode designs. The web single-mode and sequential mixed-mode designs lead to lower costs per respondent compared to the concurrent mixed-mode and single-mode paper & pencil design.

Keywords: Data Collection, Push-to-Web Strategy, Web Survey, Data Quality.

Title of abstract: Using Iterative Quick Turnaround Cognitive Interviewing to Integrate Quality Into the 2022 U.S. Economic Census

Kristin Stettler, Melissa Cidade – U.S. Census Bureau, Suitland, MD, United States

Every five years, the U.S. Census Bureau conducts the Economic Census, which provides extensive statistics about businesses that are essential to understanding the U.S. economy. Nearly 4 million business locations of all sizes, covering most industries and all geographic areas of the United States receive surveys tailored to their primary business activity. To prepare for the 2022 Economic Census, the Census Bureau consulted with a variety of experts both inside and outside of the Federal government about content. Their recommendations were evaluated, and new proposed content underwent cognitive testing to ensure the questions met Census Bureau quality standards.

Cognitive interviewing is a qualitative methodology often used by survey researchers to improve the quality of survey questions before a survey goes into the field (Willis, 2005). Given the number of topics (18), the number of questions (74), and the need to get feedback from so many different types of very specific businesses (size, industry classification, etc.), the process for setting up and conducting cognitive interviews was very complicated. We utilized a digital protocol and instrument with multiple paths to target various groups of questions to appropriate respondents, based on their industry, while also administering questions applicable to all businesses. We were able to recruit participants quickly and efficiently using an email scheduler, where respondents selected a day and time for their interview. We conducted a total of 230 phone interviews over 3 rounds of testing in 4 months. Because interviewers entered notes and quotes in real time into the electronic protocol while conducting the interview, we could export these data and quickly discern trends in question performance. This led to quick turnaround for revising questions between rounds of testing. We met with stakeholders to continue improving the questions incrementally and tested updated questions in subsequent rounds. This intensive and iterative testing strategy helped to improve the quality of the questions that will collect data for official statistics.

This presentation will describe the cognitive interviewing methodology used for 2022 Economic Census content development. We will provide examples of questions that were revised and tested in multiple iterations resulting in higher quality questions, easing respondent burden and improving the quality of reported data. We will also provide examples of topics and questions that performed poorly in testing and were eliminated from the Economic Census, improving the quality of the survey questionnaire. We will end with recommendations for improving questions for economic surveys and censuses, including effective guidelines for testing questions. We discuss effective use of targeted interviewing and iterative findings to reduce burden and measurement error and increase data quality.

Keywords: data quality; question testing; cognitive interviewing.

Title of abstract: An Enterprise Approach to Web Survey Paradata

Joanna Fane Lineback, Joseph Crimmer, Renee Ellis, John Wilen – U.S. Census Bureau, Suitland, MD, USA

In this presentation, we discuss an enterprise approach to collecting and evaluating Web survey paradata. The Web is a common mode of data collection for many federal surveys. The U.S. Census Bureau, for instance, collects data online for roughly 100 surveys and censuses. With online data collection comes the opportunity to collect a rich set of paradata, which can include information about the device and software being used to access and complete the survey, as well as information related to users' interactions with the survey instrument. These paradata outputs can be evaluated for quality assurance purposes – from improving instrument design to making real-time adaptive design changes to the data collection and evaluation process.

Nonetheless, as with many auxiliary data sources, barriers to use exist. In the case of Web survey paradata, one needs to navigate policies around data use and the unstructured nature of the data. Not all survey programs have had the need or resources to overcome these challenges. To eliminate the barriers and increase the use of this important data, the Census Bureau has been developing an enterprise approach to standardizing, parsing, aggregating, and acting upon Web survey paradata, with the goal of real-time reports for all recurring, online surveys.

Going forward, as part of the enterprise approach, all Web paradata and Web paradata reports will be housed in a centralized location, with access granted on a need-to-know basis with role-based permissions in place. Reports will be updated in real-time and will display a wide range of case level and aggregate information derived from the paradata. To mention just a handful, there will be device usage rates, operating system usage rates, login rates, breakoff rates, and error message rates. Although the reports will be as standardized as possible, the inherent nuances among surveys and survey program needs will also be reflected. We hope discussing the standardization process will serve as a useful guide to other organizations that may be considering similar approaches.

Keywords: Web paradata, standardization, unstructured data, auxiliary data.

Session 20
ASSURANCE OF QUALITY IN STATISTICAL PRODUCTION SYSTEM USING ADMINISTRATIVE, MULTI-SOURCES OR OTHER DATA SOURCES

Title of abstract: A reflection on methodological sensitivity, quality and transparency in the processing of new 'big' data sources

Fabio Ricciato – Eurostat, Luxembourg, Luxembourg

Survey and census data are primarily designed by statistical offices for statistical purposes, and for such data sources the statistical methodology encompasses both a "data design" phase (deciding what to ask to whom) and a "data processing" phase. With new "big" data sources, the data design phase is missing and the statistical methodology reduces to data processing. Due to their large volumes and rates, new data sources must be procesed automatically, and consequently the statistical methodology must be translated into software code.

Data processing methods for new data sources tend to be rather complex. Most of the new data sources tend to be generated by machines that interact with humans, and their correct interpretation requires understanding of behavioural as well as technological aspects. Furthermore, such data are often charaterised by various kinds of uncertainty and errors that must be handled properly in the data processing stage, some of which have no direct correspondance with those typical of traditional data sources. For this reason, understanding and coping with new data sources requires specialist knowledge from fields outside the traditional perimeter of statistics.

All the above means that the development of statistical methodologies for new data sources goes tightly hand in hand with the development of complex software code by multidisciplinary teams. This has several consequences for methodological development, for instances the need for methodological department to acquire skills and learns best practices from software development specialists. The "softwarization" of statistical methodology requires, among other things, the need to import estabslihed concepts, methods and practices from "software quality" into stasistical quality frameworks.

The increase of methodological complexity amplifies the need to understand, handle and communicate properly the issue of methodological sensitivity. As a matter of fact, the statistical methodology (embodied in software code) transforms huge volumes of input 'data' into a parsimonious set of output 'statistics'. The latter depends on a plethora of methodological choices that, at minimum, must be declared explicitly and documented in details in order to ensure a proper interpretation of the statistics output. Furthermore, there is a growing need to perform a more systematic analsys of sensitivity of the output (statistics) to the methodological choices taken at methodological development stage. In this respect, the "softwarization" of statistical methodologies provides new possibilities that will be elaborated in the presentation.

Keywords: Methodological quality, transparency, sensitivity.

Title of abstract: Quality Guidelines for the Acquisition and Usage of Big Data

Alexander Kowarik, Magdalena Six – Statistik Austria, Vienna, Austria

The increasing knowledge and experience within the European Statistical System (ESS) in the acquisition, processing and use of new data sources provides now a clearer picture on quality demands. These guidelines use the quality-based experiences in the ESSnet Big Data II to formulate guidelines for NSIs who already use and/or plan to use new data sources for the production of official statistics. Looking at the production process of statistics, the usage of new data sources mostly affects quality aspects of processes related to the input and the throughput phase. Taking this into account the guidelines concentrate on the input and the throughput phase of the statistical production process.

With new data sources, the access to as well as the processing of input data makes it necessary to consider new and very source- and data-specific sub-processes. The variety of sub-processes is much broader compared to the use of traditional data sources. What is relevant for one data class and one data access might be of no interest for others. We therefore decided to develop a modular approach for the structure of the quality guidelines. This allows producers to focus on the guidelines relevant for the intended form of data access and the intended data usage and takes into account the peculiarities of a specific data class.

In the paper, we will discuss the structure of the quality guidelines and some exemplary parts.

Keywords: Big Data, new data sources, quality guidelines

Relevant for the following topics:

(II) Quality Assurance in the New Emerging Data Ecosystem

– Methodological innovation and research for quality measurement and improvement
– Assurance of quality in a multi-source statistical production system
– Assurance of quality of output when using alternative data sources

Title of abstract: A path to a quality framework for a new data source: Online Job Advertisements

Fernando Reis – European Commission (Eurostat), Luxembourg, Luxembourg
Gabriele Marconi, Anca Kiss-Nagy, Raquel Paulino – Sogeti, Luxembourg, Luxembourg

In a world impregnated with smart technologies, the evolution of official statistics can be based on smart technologies that will allow the production of smart statistics that can be trusted. The Trusted Smart Statistics concept was developed as a service provided by smart systems, embedding auditable and transparent data life cycles, ensuring the validity and accuracy of the outputs, respecting data subjects' privacy and protecting confidentiality.

Following the digitalisation of our societies, new non-traditional data sources are seen as an opportunity for the production of official statistics, motivated by the multiple benefits they promise to bring, such as improved timeliness and accuracy, increased level of detail and relevance, finer temporal & spatial granularity, and low production cost. When producing official statistics, the data quality assessment is a crucial step. Currently, comprehensive analysis and research of quality standards and quality assessment methods for new data sources are rare. While working with big data, it is almost impossible to state generally applicable quality guidelines. To overcome this issue and simultaneously provide the user with clear instructions in using web data, a formulation of quality indicators is inevitable.

This paper describes data validation and model evaluation procedures to evaluate the quality of a novel data source (online job ads or OJA). Both procedures are designed as dynamic processes that combine quality assessment with improvement through cycles of feedback and revision, either of the data or the processes themselves.

OJAs refer to advertisements published on the World Wide Web revealing an employer's interest in recruiting workers with certain characteristics for performing certain work. They usually include information on job occupation and location, characteristics of the employer (e.g. economic activity) and job requirements (e.g. work experience and skills). Currently, the OJA dataset contains information on 116 851 363 distinct OJAs collected from 316 sources in Europe.

While providing important advantages compared to traditional data sources, OJA data also has limitations. Therefore, their ability to generate accurate, consistent and comparable data needs to be assessed, tested and improved. This can be addressed by implementing a data validation procedure based on specific frameworks proposed for methodology and quality with new sources of data.

Validation looks at the internal consistency of the data (and in some cases, at the consistency with external sources of data), which can be sufficient for data quality assurance in statistical applications using well-tested methods to produce the data. However, new data sources are harvested through novel algorithms that need to be evaluated. Internal data evaluation provided by the validation process, does not ensure accurate or unbiased data classification algorithms. Therefore, model evaluation procedures are to be put in place.

Keywords: official statistics; big data; web data; quality assessment; data validation.

Title of abstract: Inclusivity in Administrative Data: A Pilot Investigation of Key Population Groups

Connaire Pearce, Karina Williams – Office for National Statistics, Titchfield, United Kingdom

Administrative data is data that has been collected during the operation of an organisation, e.g., hospital records, prison records, etc. The UK Government produces a large amount of administrative data that is potentially useful for another purpose, producing statistics. However, as this purpose differs from the original intended use it was collected for, it gives rise to specific and often complex challenges for researchers.

One of these challenges is data quality. In order to use administrative data effectively and responsibly for research purposes, an understanding of the extent to which it can be considered inclusive is essential. Inclusivity gaps within administrative data are problematic as they present data quality issues when producing population statistics from admin data. These quality issues can include missingness, accuracy and the representation of vulnerable groups and groups with protected characteristics. Administrative data with these quality issues impact the potential for the data to be reliably used in planning essential social provisions and services. Although some work has been carried out to quantitatively understand inclusivity gaps in administrative data outputs, there remains a significant lack of qualitative research investigating inputs.

The Methodology and Quality Directorate at the Office for National Statistics have conducted a pilot qualitative project which focused on key population groups in order to understand how these groups interact with admin data. The groups were identified through several sources including engagement with internal topic experts. The key population groups were chosen from categories based on health, religion, migration, transient and ethnicity.

The project's methods involved interviewing key population groups' representatives, including local authorities, third sector and community organisations. The project also conducted a literature review which identified quantitative methods that can be used to overcome inclusivity gaps. A quality appraisal was carried out on the interviews and literature review's findings to assess their value and evidence base. Finally, an approach was developed to triangulate this information to provide valuable insights.

The project's findings will help the Office for National Statistics to improve the quality of the administrative based population estimates. This will be achieved by utilising insights on inclusivity and which methods could be used to resolve the gaps found. The project also provided information to the UK Statistic Authority's Inclusivity Data Taskforce which seeks to improve the UK's inclusive data holdings and admin data quality.

This presentation will cover: a) Background information; b) Methodology; c) Findings; d) Implications; e) Potential future plans (i.e., full iteration of the project).

Keywords: inclusivity; quality; administrative data; representation; key population groups.

Title of abstract: Testing for the bias in the estimation of business structure indices from different data sources

Michaela Balkoudi – Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki, Greece

Christina Karamichalakou – Business Statistics Division, Hellenic Statistical Authority, Piraeus, Greece

Dimitris Kugiumtzis – Department of Electrical and Computer Engineering, University of Thessaloniki, Thessaloniki, Greece

A key question that concerns Statistical Authorities using data collected from surveys and administrative sources for the compilation of specific statistics, is whether the usage of two different data sources (survey, administrative files) affects the computation of statistical indices. In this paper, aggregated data from the Greek Structural Business Statistics of the reference years 2014–2018 for 8 statistical indices of 140 business branches, originated from two sources, survey and administrative files, have been analyzed to assess whether the computed statistical indices from the two data sources differ. The same analysis is repeated for each of the eight statistical indices. The structure of the data is that of repeated measures (5 years) with random effects, where the within-subjects factors of the experimental design are the data source (of prime interest) and the reference year, and the random effects regards the business branches, as the level of a statistical index differs distinctly in the 140 business branches.

The statistical testing was performed using two parametric statistical tests, the two-way repeated measures ANOVA and the linear mixed models, as well as the respective bootstrap tests (using wild bootstrap for the linear mixed models). The consistency of the parametric and bootstrap tests was first assessed on simulated data, using the data setting of the real data but determining different scenarios for the dependence of the statistical index on each factor. The simulation study concluded that even for strong deviations of the data from normality the parametric and bootstrap tests agreed to the correct test decisions. The results of the real data analysis confirmed the overall agreement of the parametric and bootstrap tests, and for two statistical indices statistically significant effect of the data source was found. The results of the study suggest that the data source may have an impact on the computation of statistical indices.

Keywords: Data sources, structural business statistics, effect, repeated measures, mixed linear models, bootstrap testing.

Title of abstract: Multivariate Small Area Estimation The Case of Non-Continuous Variables

Angelo Moretti – Department of Computing and Mathematics, Centre for Advanced Computational Science, Manchester Metropolitan University, Manchester, United Kingdom

Policy makers require reliable information on the geographical distribution of spatial heterogenous social indicators at small geographical level or related to small groups of the population (small areas). However, large-scale sample surveys are not designed to produce reliable direct estimates at small area level due to the unplanned domain problem.

Univariate mixed-effect models are widely adopted in small area estimation in order to improve direct estimators by using auxiliary information. The small area estimation literature has shown that multivariate small area estimators, where correlations among response variables are taken into account, provide more efficient estimates than the traditional univariate approaches. Here, we 'borrow strengths' from both related response variables and auxiliary information.

There has been work on the use of multivariate mixed-effect models for continuous variables in both unit-level and area-level approaches. However, there are still many gaps in case of non-continuous response variables, such as binary, count or mixed-type response variables, especially in the unit-level approach.

Here, the use of multivariate generalised mixed-effect regression models is relevant and promising. Indeed, many variables in social and business surveys present a non-continuous nature and show medium/large correlations. For example, in income surveys there are many binary and count variables. In addition, crime and victimisation surveys are characterised by the presence of these types of variables, in particular count variables.

In this work, we propose the use of generalised multivariate mixed-effect models to construct small area estimators of indicators that are based on non-continuous response variables. We focus on the unit-level approach, where information is available at unit-level. A particular attention is paid on binary, count and mixed-type variables. The latter case is when, for example, one variable is binary, and the other variable is continuous. We show that there are several factors that play a role onto the efficiency of the multivariate estimators over their univariate setting (e.g., correlation structure and intra-class correlation coefficient).

In addition, we evaluate some approaches to estimate the mean squared error of the small area estimators via bootstrap. This is a flexible widely used approach to mean squared error estimation of small area estimators in presence of complex modelling strategies.

The approaches are evaluated via simulation studies and applications based on real data. In terms of application we focus on poverty, wellbeing, and crime indicators at small area level.

Keywords: Design-based; Indirect Estimators; Model-based; Random Effect; Unit-level.

NEW WAYS OF PRODUCING STATISTICS

Title of abstract: Experimental Results on European Urban Hiring Markets Using Online Job Advertisements

Andrea Ascheri, Fernando Reis, Mátyás Mészáros – Eurostat, Luxembourg, Luxembourg
Gabriele Marconi, Raquel Paulino, Anca Maria Kiss – Sogeti, Luxembourg

The recent health crisis has generated new opportunities and requests to provide new indicators to monitor the economic, social and environmental impacts of the pandemic. One of these opportunities is offered by Online Job Advertisements (OJAs) collected from several job portals through agreement, APIs and web-scraping.

We present the first experimental study using over 100 million OJAs collected from hundreds of job portals in 2019–2020 to provide evidence of the level of competition among firms (Herfindahl-Hirschman Index) in urban hiring markets across nearly all occupations and for every functional urban area (FUA) of the 27 EU Member States. The results show that across urban areas, hiring market concentration tends to be low in large urban areas in Europe (e.g. Berlin, Milan, Paris), indicating a robust degree of competition among employers and more choice for job-seekers across all occupations.

While the benefits of measuring a phenomenon of great relevance to labour markets are clear, online job ads data come with limitations, since for instance they do not represent the entire job ads population. Although this problem is likely to decrease over time as the use of web for job advertising becomes widespread, it is nevertheless an issue that has to be addressed when using OJA as well as other quality aspects such as comparability and accuracy. The implemented methodology, inspired to an idea coming from a recent European Statistical System (ESS) project, despite being new to official statistics is based on strong scientific evidence. In addition, the underlying code, developed in R, is shared and auditable on GitHub to ensure transparency and replicability.

We discuss the different quality aspects of this new data source focusing on the variables which are relevant for the calculation of the proposed indicator, such as the name of the employer, and the implemented methodological solutions. Future activities will focus on measuring accuracy and reliability both of the data sources and the statistical outputs to improve the quality of these experimental results and eventually include among European official statistics

Keywords: Experimental, Web data, Online Job Advertisements.

Title of abstract: Emerging Success Stories in the Production of Experimental and Official Statistics with Web Data

Jacek Maślankowski, Dominika Nowak – Statistics Poland

At the ESS level, the efforts to augment official statistics with web data have gone a long way. Several endeavors undertaken by Eurostat, with the substantial developments of the *ESSnet Big Data I and II* projects, demonstrated a high potential of the use of web data on the ground of experimental statistics. Concurrently, they have made the challenge of shifting from the production of experimental to official statistics even more pronounced. It has become particularly apparent in terms of ensuring a coherent system for the production of statistics with web data for all ESS members. The ESSnet project *Trusted Smart Statistics-Web Intelligence Network* (TSS-WIN), currently underway, aspires to contribute to this challenging transition.

This paper discusses the issues of the integration of web data in the official statistical production on the European scale, based on two case studies developed jointly by the TSS-WIN project consortium and Eurostat, i.e. online job advertisement (OJA) and online-based enterprise characteristics (OBEC). The work within the two domains, considered a starting point for a currently developed ESS-wide analytical environment – *Web Intelligence Hub/Platform*, provides a complex overview of the methodology-, quality-, and architecture-related issues which need to be resolved if web data is to complement and augment official statistics. The paper addresses, *inter alia,* the architecture used to collect and process the data at a central level for all EU countries, such as innovative solutions to obtain enterprises' URLs within OBEC, and data processing and validation methods within OJA.

The article concludes that while the expertise, hardware and software resources required to build ESS-wide analytical environment for the production of statistics with web data are in place, a number of legal, methodological and quality challenges need to be targeted in the first place. Varying legal regulations at national levels in terms of the transfer of data from registers, the inconsistency of the algorithms for data collection and processing, the discrepancy between the data obtained with the use of different methods are among the issues under particular consideration within the TSS-WIN project. The conclusions are followed by an overview of future plans for the development of a unified set of indicator tables for the EU countries.

Keywords: web data, trusted smart statistics, web intelligence network, web intelligence hub.

Title of abstract: From the online job advertisements to official statistics – the aspects of quality assurance

Jiri Branka, Vladimir Kvetan, Joanna Napierala – Cedefop, Thessaloniki, Greece

The abundance of existing open data and big data sources on almost any aspect of our life is creating opportunities also for production of official statistics. In 2014, the European Centre for Development of Vocational Training (Cedefop) started to produce skills intelligence using big data collected from online job advertisements (OJA) via web scraping, crawling or creation of APIs, with aim to support policy makers in designing better education and training policies. The success of this endeavor led Cedefop and Eurostat to push for the integration of big data into the production of official statistics[11].

The few years of joint project implementation brought more understanding of this data source, as well as of challenges that go with the granularity, coverage and timeliness of information obtained from online sources. These few identified challenges when building reliable time series can be grouped into four areas related to: data ingestion, de-duplication, classification of occupations and skills, and representativeness. Although in our paper we will discuss quality aspects related to all four areas of challenges, the ones linked to representativeness will be discussed in more detailed manner.

The information from online job advertisements may only be used for official statistics when three issues are addressed related to their representativeness, which are: identification of the most appropriate reference sample for OJA to assess their bias (e.g. related to coverage), estimation of the size of OJA given the observed sample in each country, and design of weighting strategy. In our paper we will present a methodology that allows for assessing the validity of collected information gathered from online sources with external data sources (e.g. Labour Force Survey, Job Vacancies Survey) based on several criteria (e.g. comparisons on occupational and geographical levels). As our data collection has started only in 2018, its length is too limited yet to allow for the testing of the time series. We will discuss the results of the pilot analysis carried out on a selection of countries (IT, ES, PL, IE). For example, the comparison between online job advertisements and job vacancy statistics indicates that certain sectors use more online websites to advertise their openings than are officially register (e.g. manufacturing sector in ES), whilst the jobs in public administration and defence were underrepresented in the online job advertisements in Ireland.

Keywords: skills intelligence; online job advertisements; quality assurance; representativeness.

---

[11] The joint project is titled "Towards the European Web Intelligence Hub – European System for Collection and Analysis of Online Job Advertisement Data (WIH-OJA)"and is carried out in the 2020-2024 period under the contract reference number:
AO/DSL/VKVET-JBRAN/WIH-OJA/002/20

Title of abstract: Italian sentiment analysis on climate change: emerging patterns from 2016 to today

Mauro Bruno, Elena Catanese, Monica Scannapieco, Luca Valentino – Istat, Rome, Italy

The debate on climate change has increasingly attracted attention, especially among young people, since the foundation of the movement Friday for Future and the raising fame of Greta Thunberg Social media websites can be used as a data source for mining public opinion on a variety of subjects including climate change. Twitter, in particular, allows for the evaluation of public opinion across time. Although it is a known problem that Twitter population is biased with respect to the whole population, it is also true that Twitter users are more likely to be young people. For this reason, the sentiment analysis of twitter textual data on climate topics provides valuable insights into the climate discussion and could be considered as representative of the rising climate movement.

In this study, a large dataset of Italian tweets (avg. 9000 daily) between 2016 and 2020 containing a set of keywords related to climate change (ex. Global warming, sustainable development) is analysed using volume analysis and text mining techniques such as topic modelling and sentiment analysis.

Topic modelling, performed by the use of word embeddings, allows validating the keywords' set and providing the major prevalent discussion in Italy about the climate agenda and their major concerns related to climate emergency.

The sentiment has been computed as a daily index by averaging the sentiment score of each tweet. The sentiment scoring has been carried out in an unsupervised setting through a pre-processed Italian lexicon, in analogy with the Italian Social Mood on Economy (SME) index.

Both daily volume and sentiment of tweets series have been analyzed. The first series allows assessing the Italian participation to the climate debate, while the latter provides useful insights on the overall evolving mood during these years. Results show that the volume of tweets linearly increased on average, and doubled over 2 years (2017-19), but since Covid the interest went back to 2017 levels, showing only in September 2021 an increase. Concerning the dynamics of the mood, Spring 2019 is characterized by optimism, while periods of extreme temperatures (summer 2017, Dec/Jan 2018, August) show a deep decrease of the index.

Daily series analysis permits to recognize special days of the Italian commitment on the debate on climate, when positive or negatives spikes are observed (such as the Global Strike in March 2019).

Keywords: sentiment analysis; climate change; social media; text mining; Twitter; Friday for future.

Title of abstract: Studiyg cyberviolence using social media data: results from an experimental statistic

Maria Giuseppina Muratore, Claudia Villante – ISTAT National Statistical Office, Italy, Rome

Violence against woman is a persistent and heavy problem that affect our society. Unfortunately, is very difficult monitoring and observing accurately the phenomenon in real time. ISTAT carried out periodical surveys but the lack of time between these surveys do not allow to collect data and information especially during specific and dramatic periods as, for example, COVID19 pandemic period. Moreover, the increased use of internet, the rapid spread of mobile information, and the widespread use of social media, coupled with the existing pandemic of violence against women, has led to the emergence of cyber-violence as a growing global problem with potentially significant economic and societal consequences. The use of Big Data in official statistics is one of the new frontiers of statistical knowledge of social phenomena. Among the various available sources of Big Data, social media constitute a particularly useful resource of information for analysing the phenomenon of gender-based violence, cyber-violence and gender stereotypes.

Through a process of capturing the contents conveyed by social networks, based on the presence of at least one word belonging to a filter set of keywords, prepared by domain experts (word embedding), it is possible to collect and observe the opinions and contents of posts, tweets and messages present on public platforms, paying specific attention to potential gender biases into the artificial intelligence algorithm used. These contents are used to apply *sentiment and emotion analysis* and to measure how the phenomenon is represented in order to verify whether social networks are used to contrast, condemn, isolate the culture of gender stereotyping or whether, on the contrary, by using a term adopted in the literature, i.e. exalting the dark side of its use, they contribute to its diffusion.

The paper, that analyse the results emerging from a specific platform aimed to capture social media content, shows which positive or negative effects in public opinion is generated, amplifying or diminishing the informative scope of certain messages. This experimental statistic is also aimed to knowing the reactions to specific information campaigns, analysing the characteristics of those who 'post' messages with the greatest numbers of followers. This study can contribute:

1) to better understand how much broaden is our knowledge of how the phenomenon of gender-based violence is perceived in public opinion and how gender stereotypes are represented;

b) to build a "thermometer" that shows which use of social networks emerges on this phenomenon (denunciation or accusation, obscurity or clarity, disgust or compassion).

This type of analysis, also known as Opinion Mining, which collects data and information directly from the web, makes it possible to add an additional piece of information to the broader information framework that ISTAT and the Equal Opportunities Department of the Presidency of the Council of Ministers have produced to date and disseminated through the integrated system https://www.istat.it/it/violenza-sulle-donne.

Keywords: gender-based violence, cyber-violence, sentiment analysis, big data.

Title of abstract: Experimental statistics on the impact of the pandemic of the corona virus disease on the Greek Economy

Adamantia Georgostathi – Hellenic Statistical Authority, Piraeus, Greece

Developments in economy, society and life, in general, are constantly changing. This dynamic process requires the National Statistical Authorities to be ready to identify emerging needs for new statistical information relevant to users' needs and opportunities to use new sources and methods, in order to provide an ever improved public good.

In the times of Covid, the need to provide up-to-date information on the impact of the confinement measures on the performance of Greek enterprises and the economy as a whole, triggered the compilation of new statistical products. Specifically, the Hellenic Statistical Authority (ELSTAT) introduced a series of experimental statistics covering different topics, important for the monitoring of the performance of the Greek economy, namely the evolution of turnover of enterprises under suspension of operation due to Covid-19, the evolution of turnover of enterprises classified in the Accommodation and Food and Beverage Service Activities and the evolution of the sales/turnover of enterprises in the activity classes of the retail trade. The experimental statistics series of ELSTAT also includes an ad-hoc publication, on the number of start-ups and closures of enterprises in Greece, providing information on the enlargement of business activity and the evolution of the economic climate-expectations in general.

The compilation of these new statistical products requires the establishment of a high frequency and sufficiently fine-tuned process of receiving and elaborating administrative tax data, in order to achieve the goal of a very timely dissemination of results, and specifically 15–20 days after the provision of the primary data by enterprises to tax authorities, each month. The standardisation of the primary tax data is performed through their connection, at enterprise level, with the Statistical Business Register of ELSTAT and administrative tax registers of enterprises by applying MicroData Linking (MDL) techniques.

The relevance, clarity and timely dissemination of these new experimental monthly statistical products of ELSTAT leaded to their establishment as acknowledged and anticipated, by users, and thus offered them a permanent position in the Press Release calendar of ELSTAT, at least for the years to come.

Keywords: Experimental statistics; Covid-19; MicroData Linking; Administrative sources.

Title of abstract: Using sensors to measure physical activity and indoor air quality: willingness, usability and data quality

Annemieke Luiten – Statistics Netherlands, Heerlen, the Netherlands

One possible element of (trusted) smart surveys is the use of fixed or wearable sensor systems. Sensors can potentially improve measurement data quality of variables that are previously measured by questionnaires. This is the case in the first example that will be discussed in this presentation: the measurement of physical activity by wearable sensors, instead of by questionnaires on physical activity, like the SQUASH. Sensors can also add insights in situations where questionnaires are hardly feasible, like in the second case study in this presentation: sensor systems to measure indoor air quality were placed in participants' homes to extract information on indoor environment that respondents can at most report subjective feelings about. Both applications would be highly informative for health statistics for example.

These promising features pose on the other hand challenges in privacy protection, trust and willingness of respondents that need careful consideration of the choice of device, recruitment message and materials. The new features also pose opportunities however, for insightful feedback to participants that may be very valuable for them and may potentially assist recruitment. During the course of the ESSnet Trusted Smart Statistics, a number of pilots has been performed in Poland, Belgium, Germany and the Netherlands, with both types of sensor systems. A wearable accelerometer was used to measure physical activity, both a high-end dedicated device, but also people's own commercial devices. To measure indoor environment quality, a commercially available sensor system was used. Both systems came with ready developed proprietary algorithms and with fully developed feedback systems. However, we have endeavoured to develop our own algorithms by employing machine learning. In the pilots, participant evaluation ascertained participant's opinions of privacy and usability. Experiments were performed with recruitment variants. In the paper and presentation, I will shortly describe the sensor systems and pilots. Subsequently, an overview of results, focussing on respondent willingness, usability and data quality will be presented.

Keywords: Smart Surveys; Usability; Pre-testing; Sensor Data; Machine Learning.

Title of abstract: Monitoring land-use in OECD cities with satellite imagery and Deep Learning

Paolo Veneri, Alexandre Banquet – OECD, Paris, France
Michiel N. Daams – University of Groningen, Groningen, The Netherlands

Over the last decades, we have been witnessing a rapid development of both space technologies and artificial intelligence (AI) applied to image processing, which have had a significant impact in various fields and have enabled to draw novel insights on the distribution of human settlements and their impact on the environment. The Copernicus Programme and the Sentinel satellites have significantly contributed to this effort. Cutting-edge technologies, such as Deep Learning, have helped to automatically draw insights from these satellite images, and to automatise tedious processes with a high accuracy, such as land-cover mapping. Such mapping is a very important tool which can have various applications: in the environmental sector, to assess the impact of wildfires, deforestation, land artificialisation, as well as in economics for urban planning and economic development purposes. It can consequently help monitor the UN Sustainable Development Goals (SDGs), and more specifically goal 11 related to sustainable cities and communities, and goal 15 concerning life on land.

This paper presents a Deep-Learning-based image segmentation framework, using the U-Net model, for a fully automatic extraction of pixel-level land-use in OECD functional urban areas (FUAs) from Sentinel-1 and -2 satellite images at a spatial resolution of 10 m. The training and validation of the model were done with a reference dataset, the Copernicus Urban Atlas, consisting of pan-European comparable land use data for 788 FUAs. The paper shows that this model enables to predict land-use in newly seen satellite images with high accuracy. It can also distinguish mostly residential, from commercial and industrial areas. It thus enables to monitor in an efficiently and timely manner land-use across European OECD FUAs. This paper also explores potential applications in non-European OECD FUAs, as well as for change detection to spot local urban expansion signals.

Keywords: Deep Learning; Earth Observation; Land-use; Functional Urban Areas.

Title of abstract: Leveraging Non-Traditional Data Sources to Enhance Passenger Mobility Statistics: Future Prospects

Andrea Ascheri, Emilia-Maria Iscru, Jean-Marc Museux, Nikolaos Roubanis, Albrecht Wirthmann – Eurostat, Luxembourg

Eurostat and DG MOVE have been exploring for some years, methods to compile statistics on passenger mobility. As no comparable information existed, Eurostat developed passenger mobility survey guidelines and defined indicators that would respond to EU policies. Since 2015, 11 EU National Statistical Institutes run dedicated passenger mobility surveys co-financed by grants from Eurostat to collect statistics, covering urban and non-urban passenger transport, short- and long-distance mobility, based on Eurostat guidelines. Active and shared mobility, new and alternative transport, means and fuels were included in the guidelines and in the dedicated surveys. Data resulting from these surveys, combined with relevant data from national mobility surveys of 3 more Member States, were published in April 2020 and were updated in 2021.Passenger mobility surveys revealed challenging in ensuring harmonised and timely results across countries with different transport patterns, despite the use of a common data collection methodology.

The availability of new data sources and privately held data has increased exponentially in recent years and the COVID-19 pandemic accelerated the transition from survey-based transport statistics towards utilising new data sources, including big data sources. Many of these data sources have helped statisticians to better capture transport and mobility pattern around the world like in the case of Google mobility reports, or AIS (Automatic Identification Systems) data used to monitor disruptions in trade and shipping.

To continue monitoring mobility across the EU, Eurostat is set to explore the use of other innovative sources and develop methods for producing statistical indicators on mobility and traffic using non-traditional data sources, which could increase harmonisation, reduce costs, burden, and improve timelines.

To exploit these new sources, Eurostat has recently started a research project jointly run by the transport unit and the methodological support team. The activities in this field started from a data exploration phase. The work will continue with assessing input quality, securing access to data and building the needed IT capabilities to store, process and assess the quality of output from these data. Finally, the methodological developments will lead to the production of experimental results and ultimately production-ready workflows. The paper will present the work in progress within this project focusing on the risks of using new data sources and on the best methods to deal with these issues.

Keywords: Experimental; Urban Passenger Mobility; Non-traditional data sources; AIS data.

Title of abstract: Data Simulation and the Modernization of Official Statistics: an example with Mobile Network Data

Sandra Barragán, David Salgado – Dept. Methodology and Development of Statistical Production, Statistics Spain (INE), Madrid, Spain
Marian Necula, Bogdan Oancea – Dept. Innovative Tools in Official Statistics, Statistics Romania (INS), Bucharest, Romania

The new merging data ecosystem, especially arising from digital data and Big Data in general, is challenging the modernization of the production of official statistics in many facets. This is especially observable in the development and incorporation of new statistical methods as well as the implications in the quality assurance framework and determination of appropriate quality indicators (e.g. model validation and accuracy assessment). Moreover, the barriers to access several new digital data sources as financial transaction data and mobile network data increase the challenge to set up a new production framework and, more concretely, to revise the quality framework to retain the high standards of traditional official statistics.

In this context, we propose to resort to agent-based models (ABMs) as a highly versatile tool to deal with many aspects of data processing and statistical production. ABMs and simulation models in general will allow NSIs to face multiple challenges at the same time:

(i) To avoid that potential legal restrictions, social concerns or economic costs regarding the access to real data, which could slow down or block methodological innovation needed to deal with these new data sources.

(ii) To develop and test different statistical methods and conduct sensitivity analyses against a known synthetic ground truth.

(iii) To develop and compute different quality indicators (e.g. accuracy) against a known synthetic ground truth.

(iv) To establish a common playground for public-private partnerships to share knowledge, designs, and plans.

We share the example of a mobile network event data simulator developed during the ESSnet on Big Data II to substantiate the so-called ESS Reference Methodological Framework. We show how a modular design for the production process is under investigation comprising the geolocation of devices, the deduplication of individuals/devices, the aggregation of individuals detected by a telecommunication network, and the statistical inference with respect to the target population.

Keywords: agent-based model, simulation, official statistics modernization, mobile network data.

Title of abstract: Experimental Statistics in Colombia, an approach for quality assessment of new data processes

Juan Oviedo, Julieth Solano, Karen Chavez, Sandra Moreno – DANE, Bogota, Colombia

In recent years, the digitalization processes have increased the amount of data collected all around the world. This increasing supply of potential sources of information for statistical use, paired with the availability of new data tools, widen the possibilities of statistical production beyond the traditional models. One of the key aspects to ensure in official statistics is quality, thus innovative production methods must be paired with quality assessment discussions. Given the lack of standards for the recent developments and fostering innovation while ensuring quality, DANE stablished criteria to define innovations and to assess some minimum quality principles of the official statistics. Currently, five measures that met the criteria have been developed with the use of non-traditional sources and/or methods, and are published as Experimental Statistics [12]. The obtained results have increased the availability of indicators to monitor the progress toward sustainable development goals, have increased the granularity of some measures previously produced, and technical capacities have been strengthened.

Keywords: quality assessment; experimental statistics; non-traditional sources; big data.

Title of abstract: Data integration: Stats Business

Cristina NEVES – Statistics Portugal, Lisbon, Portugal

The integration of survey data with administrative data has enabled Statistics Portugal (SP), under conditions of appropriate protection of statistical confidentiality, a fundamental principle of official statistics, to build a powerful enterprise database: the Stats Business.

The Stats Business, according to our best knowledge, is already one of the most advanced enterprise databases in international terms and is part of SP strategy to develop the National Data Infrastructure, with a view to taking advantage of information already available and with potential usefulness for producing official statistics. Its analytical exploration is still under development from three perspectives:

– exploitation for a more efficient production of existing official statistics;

– design of new official statistics which are made available at an early stage as experimental statistics;

– support for scientific research.

---

[12] https://www.dane.gov.co/index.php/estadisticas-por-tema/estadisticas-experimentales

With the Stats Business, SP has now available a set of information about businesses which, through microdata linking, enables the data integration of different data sources, allowing a complete perspective on enterprises' economic performance.

The most important set of information becomes from the Simplified Business Information, corresponding to an administrative source containing all the annual accounting data for all the enterprises.

SP has been developing a broader set of new statistical operations intended to disseminate information on diverse factors not covered by the European Statistical Program, but very relevant to get information on potential sources of economic growth and to capture different and important elements on business dynamics in an international context which poses new challenges and the need of constant attention to factors affecting their competitiveness: the Survey on Perspectives of Exports of Goods; the Management Practices Survey; the Survey on Framework Regulation Costs; the Survey on the Identification of Enterprises' Skills Requirements.

All this information on the enterprises, additional available administrative data (such as VAT, E-invoice system from Tax Authority, Monthly Statement of Remunerations) and other statistical data sources such as the Community Innovation Survey, International Trade in Goods Statistics, Prodcom Survey, Survey on Information and Communication Technologies Usage in Enterprises, are being used and linked at the microdata level, in order to allow the data integration and the construction of a Stats Business database. Thus, additional statistical indicators and analysis on business data are being produced, without increasing the statistical burden on enterprises.

The Stats Business database is being used already to produce experimental statistics, such as a specific "Innovative enterprises study", which illustrates the analytical possibilities associated with its exploitation, showing how innovation, activity sector, enterprise size-class and human resources qualifications are related.

Keywords: data integration; business; microdata linking; statistics; national data infrastructure.

Title of abstract: Quality reporting: where we are now and possible roads ahead

Ilcho Bechev – European Commission (Eurostat), Luxembourg, Luxembourg

In the past twenty years, supplementing statistical products and processes with reports, conveying information about their quality, has turned from a long-run strategic goal to a standard practice in all statistical domains. As a result, a critical mass of work done in the field of quality reporting standards for official statistics has given food for reflection to the practitioners. The aim of this paper is therefore to promote the existing ESS standards for quality reporting, review the recent efforts of revisiting the existing guidelines and making available best practices, ultimately consolidated in the 'ESS Handbook for Quality and Metadata Reports'. In doing so, the paper will aim also to pinpoint the possible future developments in the field of quality reporting of official statistics.

Keywords: quality reports; metadata; standards; review.

Title of abstract: Quality Reporting on National Accounts

Malgorzata Szczesna, Maria Dimitriadou, Enrico Infante, Christos Liouris – Eurostat, Luxembourg, Luxembourg

Regulation (EU) No 549/2013 establishes the European System of Accounts 2010 (ESA 2010). Article 4 of this Regulation requires Member States to provide quality reports to Eurostat, which would allow the Commission to assess the quality of data received under ESA 2010 transmissions.

Commission Implementing Regulation (EU) 2016/2304 lays down the modalities, structure, periodicity and assessment indicators of the quality reports. Under the above-mentioned regulations, Eurostat in collaboration with the Member States, has set up quality reports for National Accounts.

At the Q2018 Conference, Eurostat presented a paper that introduced the framework in which the annual quality reports are produced, the quality measures and Eurostat's efforts to automatically integrate information coming from data production into the quality reporting process.

This paper takes the topic forward and draws on the experience gathered since 2018. Among other issues, it presents new quality indicators and recent IT developments for the further automation of calculation of indicators and their introduction into the quality reports. It also highlights the data quality improvements that have been achieved by countries, in the context of the quality exercise. Lastly, it discusses the challenges that are still encountered and possible ways to further improve the process of quality reporting.

Keywords: measuring quality; quality reporting and communication; national accounts; ESA 2010.

Title of abstract: Revision Analysis of selected Austrian Economic Indicators

Mangat Manveer Kaur, Thomas Burg, Alexander Kowarik – Statistik Austria, Vienna, Austria

Economic policy implementations rely heavily on timely available economic indicators. Accordingly, the first estimates of the indicators are based on a modest fragment of the complete data set, but are revised after certain periods to accommodate the increasing availability of data over time. The analysis of the deviations between these later and preliminary published estimates, referred to as revisions, enable to draw conclusions about the reliability of the indicators. While minor deviations are to be expected and practically inevitable, the purpose of revision analysis is to assist identifying potential excessive misalignments between the estimates so that an improvement of the respective preliminary estimation procedure can be facilitated in a timely manner.

We conducted a revision analysis of selected Austrian economic indicators published by Statistics Austria, including the turnover index, the index of persons employed and the index of hours worked separately, and jointly for construction and industry. We found that regardless of whether the initial or revised estimator is considered, both exhibit a statistically significant bias. However, the size and the extent of variability of later revisions are diminutive compared to the revisions of the preliminary estimators. While the revised estimator is superior to the initial estimator for the majority of the time, the initial estimate briefly outperforms the revised estimates in certain time periods.

Keywords: Quality of Indicators; Revision Analysis; Statistical Analysis

Title of abstract: Enhanced quality reporting templates for labour market statistics on businesses

Ines Kolaković – EUROSTAT, Luxembourg, Luxembourg

Most EU Member States have already moved towards an increased use of administrative data for compilation of labour market indicators, either as a complete substitution and/or as a complement to survey data. As a result, labour market statistics transmitted to Eurostat are increasingly based on complex combinations of sources. Their quality now depends on the quality of the primary sources and the ways they are combined.

Current quality reports transmitted by Member States have been designed for survey-based data collections. Therefore, Eurostat has considered it necessary to review the content and structure of those quality reports and adapt them to this new state of play. The main driver behind this process is the extension of existing quality reports in order to accommodate new ways of data collection and allow quality assessment of administrative data and multisource data collections.

The revised quality reports are structured according to the data source: survey data, administrative source or multisource, i.e. a combination of survey and administrative data. For the latter two sources, that were not fully documented in the past, the revised template will collect information in particular on the purpose, scope and update policy of the administrative data as well as on the treatment made to adapt administrative data to the statistical concepts. In addition, more emphasis is put on checking the headline indicators against comparable data collected in other statistical domains.

For labour market statistics on businesses, this work has taken place in the context of the general ongoing review of EU legislation, discussed by a dedicated taskforce 'LMI Review'. Important progress was made and enhanced quality reporting templates for all business based labour market data collections have been prepared and will be presented.

Keywords: quality reporting; business based labour market statistics; administrative data sources.

Title of abstract: Systematic reporting on the quality of administrative data as input of the statistical production process: the Italian experience

Grazia Di Bella – Istat, Roma, Italy

The purpose of using administrative data for the production of official statistics is twofold: to reduce the burden on the respondents of direct reporting and to make available to data users more granular data. Although administrative data already represent a large share of input data for the production of official statistics in many countries and sound methodologies have already been developed to consistently incorporate those data in the production, a systematic reporting on their quality is still lacking in many countries. This paper aims at providing some insights from Istat experience on the set up and maintenance of the QRCA (Quality Report Card of Administrative Data), the Istat platform that systematically reports on the quality of Administrative Data (AD) used as input for the production of official statistics in Italy. Checking data, monitoring the acquisition processes and the pre-treatment procedures and then producing, in an efficiently and effectively manner, AD quality documentation are the tasks of the QRCA that actually manage more than 200 administrative sources.

The AD quality framework adopted, based on international best practises, considers the three hyperdimensions: Source, Metadata and Data in which quality dimensions and indicators are represented. Internal users access QRCA via intranet and can select a specific AD source through a search function. For each AD source, a variety of reports is available including structural metadata and quality indicators for any quality hyperdimension.

In Istat, the management of administrative data is in charge of a centralized structure that supports the statistical production processes. QRCA implementation is based on metadata derived from the IT tools supporting the AD management processes.

Istat published, in April 2021, the new version 2.0 of QRCA with the aim of documenting the first measures adopted for the data protection. Considering the requirements of the General Data Protection Regulation (EU) 2016/679, in the overall assessment of the quality of input data deriving from sources external to the National Statistical Institute, it is extremely important to pay particular attention to the description of the data protection strategies. Organizing and structuring relevant information in this topic is the current challenge.

Keywords: administrative data; input quality; quality report; data protection.

Session 25
MICRODATA INTEGRATION

Title of abstract: Impact of health social transfers in kind on income distribution and inequality

Estefanía Alaminos Aguilera, Sigita Geske – Eurostat, Luxembourg

Social transfers in kind are an important source of household income, especially for households at the bottom of the income distribution. The distributive impact of social transfers in kind is not explored at European level, although it has been widely studied by the international literature. Evidence of the effect of social transfers in kind is highly relevant for assessment of material wellbeing, particularly at European level before and during crisis periods.

In this work, the authors explore additional aspects of inequality not covered by existing monetary indicators (taking into account health social benefits in kind). With this purpose, the monetary income data has been complemented by imputing social transfers in kind to survey microdata (EU-SILC 2019). The health care cost (benefits) profiles by age and gender have been aligned to the government health expenditure by country provided in the National Accounts.

The first results of this work show that social transfers in kind for health, make the total income (monetary and in kind) more equal across the income quintiles. Therefore, health social transfer in kind might improve the distribution of equivalised total income across quintiles. In addition, health social transfers in kind reduce income inequalities (GINI).

These results show how the allocation of social transfers in kind in income microdata permits to carry out a deeper analysis of the income distribution, and therefore of inequalities, by different socio-economic groups of the population. The outcomes of this study could be used also in the distributional of national accounts exercise for providing methodology for the distribution of adjusted disposable income.

Keywords: social transfers in kind, material well-being, income inequalities, income distribution.

Title of abstract: Impact of data cleaning on record linkage quality

Heidi Koumarianos – Institut National de la Statistique et des Etudes Economiques, France

The French national statistical institute (Insee) aims to promote the massive and more industrialised use of administrative sources, particularly in the sphere of socio-demographic statistics. This is the context for the Répertoires Statistiques d'Individus et de Logements (RéSIL) programme, which aims to build a sustainable and evolving system of statistical registers of individuals, households and housing units, updated from various administrative sources.

Linkages between sources will be even more central to INSEE's information system.

The most automated and robust processes possible must therefore be put in place in order to achieve the desired gains in efficiency and quality.

A linkage process is broken down into several phases.

– data cleaning;
– reducing the size of the problem (blocking or indexing);
– matching of statistical units;
– comparison of units within pairs;
– classification of the selected pairs.

This paper will focus on the first of these phases, with the aim of measuring its impact on the subsequent phases.

During a matching process, the input data cleaning stage is of great importance: it ensures compliance with data formats, comparability of information, and allows matching rules to be adapted to the characteristics of the data sets.

Identified as crucial and the most time-consuming by the statisticians in charge of matching operations, it is rarely the subject of efficiency measurements.

This paper will first present the challenges of the standardisation and cleaning of individual data. This stage is based on a prior analysis of the quality of the data, notably through formal controls, or conformity to a reference system (geographical for example) and/or membership of nomenclatures.

After this theoretical reminder, different individual data sets will be linked to try to compare the impact of standardisation on data of different quality and origin (administrative data versus survey data). Several scenarios will be compared, using different levels of data cleaning: a first minimal normalisation to make comparisons possible, a more classic normalisation (deletion of non-expected characters, conformity to a geographical reference frame), and a more advanced data cleaning (exclusion of records of too low quality, more significant modification of character strings to eliminate civilities, relationships, etc.).

In each of the cases studied, we will try to measure the workload linked to normalisation and its impact on the data (proportion of modified data, excluded records, etc.).

We will then analyse its impact on the subsequent phases of the matching process.

Finally, we will try to quantify the impact of the normalisation and data cleaning on the matching result (number of records matched, exact or fuzzy, impact on the distortion of the distributions, estimates of false positives and false negatives), in relation with the chosen matching method (probabilistic versus deterministic).

Keywords: data linking, administrative data, data cleaning.

Title of abstract: Data acquisition processes improving quality of microdata at the Office for National Statistics

Verena Fermor-Dunman, Laura Parsons – Office for National Statistics, Fareham, UK

The Office for National Statistics (ONS) is the largest independent producer of statistics in the UK. Unlike many other countries, the UK has not got a central registration system and carries out a census of its population on a decennial basis. Over time information on the UK population are required more frequently and the ONS uses survey data and administrative microdata from commercial suppliers and other government departments to produce the high-quality statistics needed. Survey data can be expensive and burdensome for respondents, therefore the ONS aims to make the best use of administrative microdata available (Data Collection Transformation). Acquiring administrative microdata comes with many challenges, such as a lack of supplier engagement and transferring large amounts of microdata across different platforms, while respecting relevant data protection legislations, such as General Data Protection Legislation (GDPR).

This paper presents how ONS manages the administrative microdata acquisition process from early data source research to ongoing management by introducing a 7-stage model. This model helps ONS identify and take care of issues promptly during acquisition, ingest and management of administrative microdata. The model contributes to the improvement of microdata quality and statistics.

Successful examples of overcoming challenges using the 7-stage model and acquiring large high-quality microdata datasets includes the delivery of Council Tax data (Council Tax is a compulsory tax on domestic property in the UK) and Vehicle and Driver data, which were both used for quality assurance of the 2021 Census in England and Wales. Both data sources were providing highly valuable information not previously available to ONS.

For Council Tax data acquisition, 300+ individual datasets had to be acquired, resulting in having to overcome challenges that could have impacted on the quality of the microdata in terms of timeliness, accuracy, consistency and comparability. In turn, the acquisition of Vehicle and Driver microdata was challenging as it took place during Covid-19 restrictions which resulted in a change of supplier priorities and reduced resources. This could have impacted the quality of the data we were acquiring in terms of relevance, accessibility and timeliness.

The 7-stage acquisition process enabled us to recognise and address challenges early on and acquire high quality microdata that were timely, relevant and as accurate as possible. Furthermore, data were easily accessible by our data users and comparable and linkable to other sources. In short, the 7-stage process allows us to acquire microdata of the highest possible quality with quality being defined by the five quality dimensions outlined by the European Statistics Code of Practice.

Keywords: Data management, Acquisition processes, Microdata quality, Administrative data

Session 26
CENSUSES IN THE 21ST CENTURY [1]

Title of abstract: Quality of register-based Census results: Latent Class Analysis application

Kornelija Janušytė, Augustė Skrebutėnaitė – Statistics Lithuania, Vilnius, Lithuania

Statistical information is usually based on several data sources with different information. The increased availability of a large amount of administrative data at the National Statistics Offices makes it necessary to investigate new methods for selecting relevant information and producing accurate multi-source statistics. The main purposes of administrative data aren't just statistical information publication while administrative data providers knowledge about data quality can contravene Statistics Office data quality aspirations. 2021 is the first year for Statistics Lithuania using only administrative data for conducting Population and Housing Census, so the quality of input and output data must be considered before output dissemination. Input data quality control process is centralized and automatized in Statistics Lithuania, so our future aim is to make a system for output quality control. Until now, Statistics Lithuania was controlling only input data quality and didn't compare administrative data based on modelling. However, some population characteristics appear in several sources which information can vary based on differences between the definitions of variables, information timeliness in separate sources, human errors etc. Paul P. Biemer (2004) presented the use of latent class analysis for obtaining estimates of inconsistency and reliability. ESSnet KOMUSO introduce Latent Class Models as one of the possible quality measures and calculation methods. ISTAT also uses Latent Class Models for measurement error as a valid method for estimating measurement errors when the data of interest are categorical data. These findings show that latent variable application can help to decide which source gathers the closest information to the "true", error free phenomenon. Based on these examples, the mentioned model is applied to the analysis of Lithuania's administrative data of population.

Keywords: administrative data; Population and Housing Census data; multisource statistics; Latent Class Models.

Title of abstract: Challenges and prospects of the Hellenic Statistical System 2021 population and housing census and beyond

Vassiliki Kyprioti Benaki – Helenic Statistical authority (ELSTAT), Athens – Piraeus, Greece

The 2021 Population and Housing Census in Greece was planned to be conducted following the traditional method based on the direct count of all individuals and the collection of information on their characteristics with paper questionnaire through interview with the respondent. Introducing another method such as the combined census model in Greece was not straightforward for several reasons. The main reasons were the lack of suitable registers at national level and the lack of a unique ID for both citizens and dwelling that makes a burdensome undertaking any linkage between registers while the strict data protection regulations among different administrations presented an extra challenge to this end.

In the context of the 2021 General Censuses ELSTAT introduced a special Law governing the General Censuses voted by the vast majority of the members of the Hellenic Parliament, gaining broad support from both government and opposition parties, without any substantial amendments to the legal act proposal as introduced by ELSTAT. It is the first time that the General Censuses in Greece are conducted by virtue of a legal act that enjoys democratic legitimization to that extent.

The strengthening of the statistical legislation shows that the professional independence of the ELSS is widely recognized as a fact that provided a solid basis for the ELSS to be able to meet both current and future challenges. In particular, the institutional arrangements ensured the smooth conduct of the censuses in accordance with the international recommendations and standards, while at the same time ensuring the immediate response of the system to current and future developments.

In this context, the ELSS was able to respond immediately to the conditions arising from the Covid 19 pandemic crisis adapting the original design with an on line census of all households with special care taken to support those who are not able to complete the census questionnaire on line while the foundations for the establishment and maintenance of Statistical Registers for Population and Dwellings have been laid introducing a unique ID and link them with registers from other data sources ensuring the indispensable collaboration and support of other public agencies.

Keywords: census; registers: unique ID; statistical legislation.

Title of abstract: Quality Assurance Plan in Albanian Census 2022

Alma Kondi, Helda Curma – Institute of Statistics, Tirana, Albania

The Population and Housing Census (PHC) in Albania is an essential source of precise count of the residential population and is the most expensive statistical operation that Institute of Statistics (INSTAT) may carry out. Indeed, the planning of the future deeply depends on the availability of accurate information about what we are, who we are and how we live.

Establishing a quality assurance plan is crucial task to the success of the overall census operation. The plan should be established at all phases of census operations, including planning, pre-enumeration, enumeration, document flow, coding, data capture, editing, tabulation and data dissemination. This plan should be an instrument assess and measure quality in the overall census process.

INSTAT defines quality within 11 dimensions and they are overlapping and interrelated. However, there is no effective model for bringing together all of these characteristics of quality into a single indicator. Every dimension has to be adequately managed if information is to be fit for use; failure in any one dimension will impair or destroy the usefulness of the information and they all recognize not as a single measure of data quality.

This paper will present an overview of the quality assurance plan in INSTAT and the necessary information that are needed to manage effectively census quality from: planning to dissemination of results. It defines the quality assurance measures, quality control measures and corrective actions: that are needed to resolve anomalies and problems.

Keywords: census, quality, assurance, dimensions, action.

Title of abstract: Quality assurance of EU population and housing census outputs

David Thorogood, Fabian Bach, Fabio Sartori, Ani Todorova – Eurostat, European Commission

Population and housing censuses are a key pillar of official statistics, providing geographically detailed multi-variate statistics about the population and its demographic, socio-economic, migratory, family, household and housing characteristics.

The 2021 population and housing censuses in the EU are ongoing and expected to ensure statistical outputs at EU level in 2023 and 2024. Although the methods and data sources used vary greatly between countries and are subject to major changes, the 2021 EU census programme defines a large set of harmonised data and microdata outputs that are supplied by the EU and EFTA countries.

The paper will review the requirements for transmission of data, metadata and quality reports from Member States to Eurostat and discuss how data quality is ensured at European level in line with the quality criteria for European statistics. It will also address how users are appropriately informed about sources and methods applied as well as any particularities of the data.

A new development for the 2021 censuses will be the dissemination of some data at the level of a standard 1km square grid. This type of geo-referenced output allows for new types of data analyses to be undertaken but also poses additional quality assurance issues.

The paper will also provide an outlook for the future censuses in the context of the Commission initiative for redevelopment of European statistics on population and highlight expected quality improvements in population statistics.

Keywords: 2021 population and housing censuses; quality assurance; future of censuses in the EU.

Title of abstract: The methodological and technological innovations of data collection for the 7th Census of agriculture: process quality

Loredana De Gaetano, Claudia Fabi, Vincenzo Triolo – ISTAT, Rome, Italy

The strategy of the Agricultural Census 2021 was based on an integrated data collection system, entirely on digital support, which made it possible to use three different survey techniques simultaneously: CATI (Computer Assisted Telephone Interviewing), in both "inbound" and "outbound" meanings, CAWI (Computer Assisted Web Interviewing) and CAPI (Computer Assisted Personal Interviewing).

Istat have made available to the CAPI survey network (Agricultural assistance centers), the control network (Regions) and the agricultural companies (responding units) an online acquisition system, that has allowed the management of the multi-channel technique (CAWI, CATI and CAPI) through an IT architecture, asynchronously interfaced with outsourced CATI systems Furthermore, for the CAPI survey network, the control network and the Contact Center service to support respondents, a web application (Survey Management System – SGI) has been made available, integrated with the digital acquisition system, able to bring together in a single repository microdata and metadata coming from various survey techniques used in data collection process and to manage and monitor the phases of the survey process, with purposes consistent with the roles of the various subjects involved.

This work will analyse the three data collection techniques used for the first time in the Agricultural Census (CATI, CAWI and CAPI), the innovative tools to support the survey (SMS, WhatsApp) and the new technological solutions that allow the validation of the qualitative and quantitative trend of the multi-technical survey in real time.

The proposed analysis aims to make possible a comparison between the three data collection techniques and the IT tools, used for the first time in the Agriculture Census, also in view of an in-depth analysis and optimization of the synergy between data collection techniques and the tools developed in other surveys to be able to adopt them, in order to improve the quality of the data collected and the quality of the survey process, to respect EU constraints and to reduce statistical burden on respondents.

Starting from the comparison of data collection techniques and the analysis of the innovative technologies used in the Agriculture Census, the effectiveness of the operational and technological integration achieved in the information collection process will be assessed in order to design a set of tools oriented towards the efficiency, effectiveness and quality of data collection, also from the perspective of the permanent census and all other surveys, including sampling ones, which will allow the use of a data collection strategy in synchronous multi-technique mode.

Keywords: data collection, multi-technique, innovation, integrated system, process quality.

Title of abstract: Population Census 2021 in the Slovak Republic: The "Signs of Life" method

Mgr. Kristián Óvári – Statistical Office of the Slovak Republic, Bratislava

The 2021 Population Census in Slovakia was based on combining data from administrative data sources and data from field enumeration through electronic forms. The implemented concept determined processing the population census results. Every choice of a given administrative source had to be preceded by a data quality assessment, which determined the importance and relevance of each register used for getting a particular information. Data collection results were merged to an integrated database (IDB), where all the data had been linked to each other by national identification numbers. Each person living in the country leaves certain traces of activity, which are recorded in one or more registers.

The so-called „Signs of Life" method provide a set of combination of the mentioned registers for each person, which served as a proper tool for assessing the population size of the Slovak Republic. The „Signs of Life" method was applied on the records of the IDB and subsequently the number of inhabitants of the Slovak Republic was determined according to permanent, current and usual residence. The aims of this paper is to describe the process and the results of the „Signs of Life" method, which was applied in the case of the Slovak Republic.

Keywords: Signs of Life; Integrated Database; administrative data source; Population Census 2021.

Title of abstract: Non-probability sample integration in the survey of Lithuanian census

Andrius Čiginas – Statistics Lithuania, Vilnius University, Vilnius, Lithuania
Ieva Burakauskaitė – Statistics Lithuania, Vilnius, Lithuania

The sample of the Statistical survey on population by ethnicity, native language and religion 2021 consists of the voluntary sample and the probability sample drawn from the rest of the census population. A natural post-stratified calibrated estimator tends to underestimate minor religions and other small proportions of interest. Alternatively, to correct the bias of estimates based on the non-probability sample, we evaluate the propensity scores for individuals using the demographic and the previous census data. We show that the combination of the corrected estimates with the calibrated ones improves the estimation accuracy for minor religions.

Keywords: population census; missing at random; propensity score; variance estimation.

Title of abstract: Evaluation of quality of item imputation in census data under varying missingness mechanisms and non-response levels

Olga Krikun – Office for National Statistics, United Kingdom

A population census is the largest statistical operation undertaken every 10 years by the Office for National Statistics in England and Wales. Information collected during the national census informs an array of policies. Irrespective of how well the census processes were implemented, the collected data are rarely without errors or missingness. Statistical data editing and imputation aimed at correcting inconsistencies and item missingness is an integral part of census operations.

This presentation is based on research that seeks to expand the evidence base around the performance of the item imputation approach in the context of the England and Wales Census data. The main goal of the research paper is to explore properties of the chosen imputation method with a particular focus on imputation accuracy and uncertainty. The simulation study conducted assesses the performance of imputation under varying non-response levels and the two missingness mechanisms: missing completely at random (MCAR) and missingness at random (MAR). The main contribution of this study to the existing body of research is in the proposed approach to generating multivariate missingness with dependencies via the application of log-linear modelling to contingency tables.

The simulation study confirmed that the nearest-neighbour imputation methodology is sufficiently robust to impute multiple variables with varying levels of item non-response and missingness mechanisms. In the context of item non-response observed in the 2011 Census both bias and imputation variance are very low. Overall, variance is not substantially different under varying missingness mechanisms. However, the level of bias looks more concerning as item non-response increases. Moreover, bias is substantially higher in instances when MAR assumptions are employed. There appears to be evidence that the success of recovery of the true distribution of data categories are directly linked to the size of the donor pool, the availability of information on matching variables and the underlying composition of households.

The high non-response scenarios implemented in this study under MAR assumption are quite extreme, as in some instances up to a third of a population in a given category was made missing. It is unlikely that census data would be affected by such a sizeable item missingness. However, in survey imputation non-response levels can be significantly higher than what was modelled as an extreme scenario in this study. A similar approach for generating multivariate missingness under MAR assumption can be adopted to assess the extent of error in wider applications, such as surveys or administrative sources.

Keywords: item imputation; missingness mechanism; log-linear modelling; contingency tables; simulation study.

Session 28
CHALLENGES IN THE HOUSEHOLD SURVEYS

Title of abstract: European social indicators' revision and modernization: the special case of the AROPE indicator and its components

Estefanía Alaminos Aguilera, Emilio Di Meglio – Eurostat, Luxembourg

The at risk of poverty or social exclusion indicator (AROPE) corresponds to the sum of persons who are either at risk of poverty, or severely materially deprived or living in a household with a very low work intensity.

AROPE has been the headline indicator to monitor the EU 2020 Strategy poverty target and remains crucial to monitor the EU 2030 target on poverty and social exclusion. The EU 2030 poverty objective aims to reduce by at least 15 million the number of people at risk of poverty or social exclusion by 2030, out of them, at least 5 million should be children.

This work shows the decisions that have been taken by Eurostat in strict cooperation with policy users to modernize two of the components of AROPE. Firstly, the changes to adjust the severe material deprivation indicator, based on some obsolete items towards the severe material and social deprivation measure will be shown. Secondly, the changes adopted regarding some of the parameters of the low work intensity indicator, to adapt it to the new realities in the labour market will be described. Finally, the impact of these changes on AROPE will be shown.

The revision entails a marginal impact on both the levels and composition of the overall population considered at risk of poverty and exclusion at EU level. This will allow to capture a broader and more refined measurement of deprivation, based on a revised and elaborated list of items, as well as to better account for the social exclusion situation of those of working age from age 18 to age 64.

Keywords: social indicators; poverty indicators; at-risk-of-poverty or social exclusion indicator (AROPE); deprivation; social policy monitoring.

Title of abstract: Stratification criteria based on income data from administrative registers for the first wave samples of EU-SILC in Austria

Thomas Glaser, Johannes Gussenbauer, Alexander Kowarik – Statistics Austria, Vienna, Austria

The yearly survey EU-SILC (European Community Statistics on Income and Living Conditions) in Austria uses income register data from administrative sources. These data are linked to the sample of EU-SILC on micro-level to survey about 87% of the volume of the total household income. Since this information is also available for the sampling frame, it can be applied in the sampling design. The goal is to improve the quality of results by reducing the standard error of the main indicators on poverty and social exclusion in EU-SILC. These indicators are mostly based on income information from administrative registers: rate of people at risk of poverty or social exclusion (AROPE) and the at-risk-of-poverty rate (AROP).

From EU-SILC 2016 onwards, stratification of the first wave sample of EU SILC was carried out by applying a sub-stratification on regional level, which used income information from administrative registers in the sampling frame. For each person registered at addresses of private households in the sampling frame of EU-SILC 2016 and 2017, income data from registers were matched. The sum of all net income components was then aggregated for each address in the sampling frame, creating a net household income based solely on register information. The first quartile of this characteristic is correlated with AROPE and AROP. Therefore, it was used as a sub-stratification criterion within regional strata for the selection of the first wave sample of EU-SILC 2016 and 2017.

For the first wave of EU-SILC 2018 to 2021, this approach was enhanced by using the newly available "rich frame". This enhanced sampling frame is a quarterly generated frame of the whole Austrian population based on several registers. Sociodemographic variables in combination with the available income information were used to train a machine learning model (random forest) to estimate the AROPE for the entire frame. This predicted AROPE was then used as a sub-stratification criterion within the regional stratification by Austrian provinces.

The effects of the abovementioned sub-stratification criterion on the standard error of AROPE and AROP are compared. Results show a tendency towards a lower coefficient of variation of AROPE and AROP if sub-stratification is applied. If the standard error estimation does not take into account the effect of the calibration of survey weights, the observed reduction becomes more apparent.

Keywords: sampling, rich frame, standard error, administrative registers, EU-SILC.

Title of abstract: Shedding light into the black box of missing income values: Item nonresponse and imputation practices for income variables in EU-SILC

Richard Heuberger – Statistics Austria, Vienna, Austria

Item nonresponse and imputation are crucial topics in social surveys data users are seldom aware of. They become even more important when asking for income variables and, in particular, if the survey aims at cross-country comparisons like the European project European Union Statistics on Income and Living Conditions (EU-SILC). Examining item nonresponse and imputation practices of income variables in more detail was the goal of our contribution to the research project Net-SILC 3, which has been accompanying EU-SILC since 2017.

Starting with the analysis of causes and mechanisms of item nonresponse, the variety of different preconditions for data collection in European countries participating in EU-SILC became visible. Flag variables are included in the datasets of EU-SILC as a tool in order to highlight the thereof resulting cross-country differences in item nonresponse and imputation practices to the data users. However, as our research points out, the shortcomings of the flag variables make it difficult to consider the different mechanisms of item nonresponse and approaches to imputation.

Despite the encountered difficulties in working with flag variables, in the next step of our research we analysed cross-sectional and longitudinal imputation practices in EU-SILC. Therefore, we empirically tested the effect of different cross-sectional imputation methods on the resulting income distributions before and after imputation. Additionally, in a simulation study for longitudinal imputation practices, we highlighted the importance of the decision for one or the other imputation method. A specific approach to income imputation is gross/net conversion, which is of particular importance for an income-centred survey like EU-SILC and, thus, was also paid attention to in the Net-SILC 3 project. Lastly, we formulated suggestions for improvements for the ongoing revision of EU-SILC, which should, at best, be addressed by NSIs and EUROSTAT similarly.

Keywords: EU-SILC, item nonresponse, imputation.

Title of abstract: How to optimize data quality in a smart household budget survey?

Jelmer de Groot, Barry Schouten – Statistics Netherlands, The Hague, the Netherlands

Household budget/expenditure surveys contain many elements that make them very well fit for the introduction of smart survey features. These surveys tend to be burdensome, both in time and in cognitive effort, and the information requested may not always be readily available for respondents. Smart surveys introduce features of smart devices such as internal storage and computing, internal sensors, linkage to external sensor systems, access to public online data and various forms of data donation. Some of these features, such as receipt scanning or uploading, advanced product search algorithms and data donation, are very promising. They are also challenging in terms of user interface design and processing through text/image extraction methods.

In ESSnet Smart Surveys a large-scale field test has been conducted testing various recruitment and motivation strategies. Randomized conditions such as mode of invitation, in-app feedback on OCR results and personalized insights were added. These conditions allow for an analysis of data quality trade-offs. Since the aim is to reduce perceived respondent burden, it is crucial to know what the boundaries are in respondent involvement, and how these depend on design features.

In the paper, we will discuss data quality for a smart household budget survey as a function of open design choices.

Keywords: Smart Surveys; Usability; Pre-testing; Sensor Data; Machine Learning.

Session 29
FOCUS ON USERS TO INCREASE THE VALUE OF OFFICIAL STATISTICS

Title of abstract: Enhancing the value of official statistics: conceptual foundations

Andrea Ordaz-Németh – Hungarian Central Statistical Office, Budapest, Hungary

In an increasingly saturated data market, it is no longer enough to insinuate the value of official statistics, it must be proven. There is an underlying expectation that by better communicating quality frameworks and core values to users, they will deduce the importance of official statistics and National Statistical Offices (NSOs) will thus stay ahead in the race with other competitive data actors. The core values of official statistics function as an anchor for NSOs. However, the value of official statistics as a whole, shifts in response to evolving demands. Therefore, communicating our core values does not deliver the entire message of the differences between official statistics and third-party statistics to users. Similarly, other players in the market may follow quality guidelines which may be close enough to NSOs standards to make little difference to users. As producers of official statistics, NSOs can go one step further and articulate the value they provide to users in relation to the concepts of quality and core values. To this end, the paper proposes a conceptual framework that relates the core values and quality frameworks of NSOs to the value of official statistics, and on which a model can be developed for the monitoring of the user-perspective value of official statistics. This approach builds on the work being carried out by the UNECE Task Force to test and develop the framework for measuring the value of official statistics.

The COVID-19 pandemic highlighted the value of official statistics as the basis for policy changes. NSOs are presented with an opportunity to demonstrate to users the ways in which they uphold their commitment to core values and high quality while adjusting to new realities in order to enhance the value of official statistics.

Keywords: value of official statistics; core values; user-perspective; conceptual framework.

Title of abstract: User satisfaction survey of sub-municipal data from the 2011 Italian Population and Housing Census

Giancarlo Carbonetti, Roberta Roncati, Antonella Ciccarese – ISTAT, Rome, Italy

Since 2015 and following the dissemination of the results of the 15th Italian Census of Population and Housing (2011), many requests for customized data processing by external users have been received by Istat Contact Center, the institutional channel to request data. These requests referred almost exclusively to data not designed in the Italian Dissemination Plan and related to sub-municipal domains (census area; enumeration area).

Appropriate methods and tools have been developed to meet these demands. The data requested mainly referred to educational attainment of population, current activity status, status in employment, occupation and branch of economic activity of employed, types of households and structural characteristics of dwellings.

Over the years, some questions have arisen about the data provided:

– did they fully satisfy the users?
– did they allow to achieve the set study, research, analysis and planning objectives?
– were they used for other purposes not initially intended?
– did the level of detail and quality of the data provided really satisfy the user?
– were the data also used through the use of GIS?
– has the data received been integrated with other sources available to the user?
– will the user in the future need to ask for new, more timely and more detailed data from a classification or territorial point of view in the continuity of their activities?

This last question is related to the important opportunities that the current Permanent Census of Population and Housing in Italy will have to offer reliable data at the sub-municipal level; there is a great expectation from users of highly detailed and timely spatial data every year.

To answer the questions listed above, a survey was carried out on the external users who in recent years have requested customized processing of data from the 2011 Population and Housing Census referring to sub-municipal domains and not available on the Istat web dissemination platforms. This survey aims to evaluate:

– relevance and satisfaction of the data provided by Istat;
– achievement of the purposes of the data request and any other objectives;
– possibility of future requests for similar data or with more details.

The high response rate obtained has favoured the collection of useful information from users. The data analysis will allow us to identify the information needs of census data by external users, also in the perspective provided by the Permanent Census which renews its data offer from year to year; moreover, it will allow us to evaluate

the level of specialization of users who request data with a high informative content. Finally, it will be possible to draw useful information for the different Istat structures that contribute to the census operations: from the planning of the data collection phase, to the definition of the thematic registers, from the development of estimation methodologies for small areas, to quality issues, up to the different methods of disseminating data.

Keywords: users needs; user satisfaction; sub-municipal area; dissemination; census.

Title of abstract: Statistical confidentiality for tabular data in European Integrated Farm Statistics

Denisa Camelia Florescu, Elio Litti, Helena Ramos, Petronela Reinecke – Eurostat, Luxembourg, Luxembourg

The paper addresses the question: "How could the utility of statistics increase for the users without compromising the effectiveness of the confidentiality treatment?"

Under Eurostat's coordination, the statistical bodies belonging to the European Statistical System (ESS) are collecting data at the level of agricultural holdings on the structure of farms, production methods, rural development measures, agro-environmental aspects and other related information. Eurostat disseminates a large number of statistical tables on its website and through specific user requests. All these tabular data are treated for primary confidentiality. The methods used are cell suppression due to small counts and due to dominance by two units as well as rounding of values.

This paper outlines the latest improvements to the confidentiality treatment, the current situation and discusses possible future options, reviewing their pros and cons. Numeric examples illustrate the share of suppressed cells in tables with various dimensions.

The objective is to strike a good balance between the amount and value of statistical data made available to users and the protection of individual observations.

Keywords: integrated farm statistics; farm structure surveys; agricultural census; confidentiality; suppression.

Title of abstract: The Art of Formulating Quality Requirements for Official Statistics

Heather Bergdahl – Statistics Sweden, Örebro, Sweden
Henrik Petterson – Transport Analysis, Stockholm, Sweden
Linn Christiansen – Swedish Forest Agency, Jönköping, Sweden

Formulating general requirements for quality in official statistics is difficult as requirements vary between different user needs and different types of statistics. A reasonable level of accuracy may vary between study domains with different levels of detail. Certainly, official statistics should be punctual, but timeliness, on the other hand, will depend on the content, data collection method, etc.

One type of general requirement could be to set a maximum acceptable level for the nonresponse rate for a given level of quality. However, even this proves to be difficult if not impossible. The important aspect here, is the effects of nonresponse on the estimates which are dependent on response rates but also, for example, on the representativity of responses.

The Swedish quality concept for official statistics starts with the purpose of the statistics which is formulated in dialog with the users. It states which information needs will be satisfied by the statistical survey and its output. Quality is thereby set in relation to the intended uses indicating its "fitness for purpose".

From the purpose of the statistics, which is survey-specific, follow quality requirements which are formulated, advisably in dialogue with the users, and decided by the responsible agency. The quality requirements determine the minimum acceptable levels for each of the different quality components (Relevance, Accuracy, Timeliness and punctuality, etc). Formulating quality requirements is subsequently a crucial step in forming the design of the statistical survey and later its production.

How do quality requirements follow from the purpose of the statistics? Is it possible to assess if the stated quality requirements are fulfilled and if the statistics are fit for their purpose? In this paper three of the statistical agencies in Sweden will explore the answers to these questions.

Keywords: quality requirements; purpose of the statistics; user dialogue; fitness for purpose.

Title of abstract: Statistical notes as a strategy to accelerate the effective use of official statistics in Colombia's Data ecosystem

Juan Oviedo, Julieth Solano, Karen Chavez, Karen García, Mauricio Ortiz, Brandon Rojas – DANE, Bogota, Colombia

With the aim to foster the use of data as a decision-making tool, DANE created the Statistical Notes Series, as an innovative strategy for dissemination of statistics. The value of data as a public good increases in its use for evidence-based analysis, since it improves the understanding of the social, economic and environmental phenomena of the country, allowing better policy designs. The Statistical Notes are documents that bring together all statistical information available, from DANE and other entities of the NSS, to characterize a topic of interest. It incorporates visual proposals from the cognitive perspective to improve the understanding of the contents. As part of the dissemination strategy, the following are included: the organization of panel discussions with various actors, and the creation of infographics pieces for social media. After the publication of 16 Statistical Notes [13], there is an increase in the supply of information available, the number of consultations, the types of data users, and a greater articulation with other actors in the data ecosystem. To this extent, Statistical Notes contribute in making data more timely, relevant, accessible and consistent.

Keywords: value of data; dissemination; visual proposals; accessibility.

Title of abstract: Quality, values and value: three essential but different aspects of official statistics

Angela Potter – Office for National Statistics, Newport, United Kingdom
Giles Sullivan – Independent expert, New Zealand
Fiona Willis-Núñez – United Nations Economic Commission for Europe, Geneva, Switzerland

A company is building a bridge.

They break the job down into sub-processes, tools and materials, each associated with quality control metrics. They check the quality of every component and every task before, during and after construction. They build a beautiful bridge.

But the bridge is no good. It is sturdy and elegant, but it does not connect two locations that needed connecting. Nobody needs to cross the river at that point, so the bridge remains unused.

High quality components and processes, and even a high-quality finished product, don't automatically lead to a valuable result.

---

[13] https://www.dane.gov.co/index.php/servicios-al-ciudadano/servicios-informacion/serie-notas-estadisticas

In official statistics, too, high quality processes, products and services are no guarantee of value to the customer. This paper, inspired by a Conference of European Statisticians (CES) task force on Measuring the Value of Official Statistics, argues for and explains the differentiation between *quality*—how good things are—; *values*—the things that drive and motivate us—; and *value*—what our work is worth, its usefulness in the eyes of our users and for society.

A clear understanding within the official statistics community of the meaning and distinction between these three features is an essential precursor to communicating with society about all three concepts.

The presentation will introduce the central arguments underlying the CES task force's work, which has recently been subject to consultation among all CES member countries and which is due to be endorsed at the forthcoming CES plenary session. It is hoped that the presentation will stimulate much-needed introspection and debate about the importance of separating out these three crucial concepts which are so often subject to confusion, not only for linguistic reasons but for more fundamental reasons related to the way that official statistical organizations tend to operate.

Keywords: quality; values; value; Fundamental Principles of Official Statistics.

Session 30
PAVING THE WAY TO STATISTICAL LITERACY/UNDERSTANDING THE WORLD THROUGH DATA

Title of abstract: Statistics Coded – How literate programming can pave the way to statistical literacy

Mátyás MÉSZÁROS, Jean-Marc MUSEUX and Jacopo GRAZZINI – Eurostat, Luxembourg, Luxembourg

The production of knowledge lies at the heart of modern organisations. However, the value of any knowledge product hangs on its effective dissemination to present and future audiences. In the light of the recent debate around *post-truth society*, the need for immediate action to further improve the dissemination of Official Statistics to wide and diverse audiences is evident.

The growing need for a more interactive dissemination of Official Statistics is accompanied by a substantial gap in the current communication system, which has in most cases an overarching top-down ideology. An opposite to it, inspired by the reproducibility movement in *Open Science*, we explore an innovative bottom-up approach to sharing and dissemination that enables collaboration and increases participative forms of statistical services' design and sharing. This approach is rooted in principles of participation and transparency, enabling others to collaborate in, contribute to, scrutinise and re-use statistical products and spread knowledge as widely as possible.

We followed the approach of Open Science, creating computational narratives of how data are collected and algorithms are constructed, to produce basic descriptive statistics based on Eurostat public data. Practically, we are accompanying official statistics with interactive computing documents, the so-called *Statistics Coded*, that interleave explanatory text with executable code and simple visualisations.

The *Statistics Coded* offer a narrative of any given statistical analysis that is both human and machine-readable. The interactive nature of the *Statistics Coded* makes it quick and easy to try out and compare different approaches or parameters. Users can fully reproduce experiments, by rerunning or tweaking previous data analyses, they can also judge for themselves. Advanced users can mine open data to explore patterns or discover problems. This way, the *Statistics Coded* also help sharing best practices, learning from each other's experience, and adopt common methodologies.

The paper will conclude with some lessons learned from this experimental work.

Keywords: Statistical literacy & literate programming; Data analytics dashboards and notebooks; Reusability and reproducibility; Public engagement & participation; Quality & trust.

Title of abstract: Statistical literacy in the OECD Recommendation on Good Statistical Practice and other international guidelines – monitoring and ways forward

Adrian Zerbe, Julien Dupont – OECD, Paris, France

International statistics guidelines (UN Fundamental Principles for Official Statistics, European Statistics Code of Practice, UNECE Generic Law on Official Statistics, OECD Recommendation on Good Statistical Practice, ISI Ethics Declaration...) emphasize the role of national statistical authorities to comment on erroneous interpretation and misuse of statistics and to enhance trust in official statistics. However, guidance to promote statistical literacy, and related data literacy, among users with the specific aim to avoid official statistics misinterpretation remains limited. Considering the expanding scope and complexity of statistics in the day-to-day life of citizens, the ability of users to identify, understand, analyse, evaluate and communicate statistical information is becoming increasingly important and the role of official statistics in adapting to this challenge is crucial. In setting standards for sound and credible national statistical systems, international guidelines need to constantly innovate in developing products and services for statistical education and communication.

This paper analyses some relevant examples of good practices developed by national statistical authorities in the area of statistical literacy, in particular across Adherents to the OECD Recommendation on Good Statistical Practice. Based on this examination, it will explore the definition of statistical literacy in international standards, and identify a set of relevant good practices. To this end, it is important to identify various user's and producer's needs and skills to decide on measures to promote statistical literacy.

Keywords: statistical literacy, international guidelines, OECD Recommendation on Good Statistical Practice, data-driven world, data literacy.

Title of abstract: Framework of presenting climate change Data for Statistical Purposes (Egypt Case Study)

Dr. Mostafa Mohamed Salah – Central Agency for Public Mobilization and Statistics (CAPMAS), Cairo, Egypt

Environmental data and information, like the natural, climate change and urban environments they describe are very precious resources. The importance and value of access to high-quality environmental data is almost universally understood and accepted.

Today, there are growing numbers of stakeholders demanding access to environmental data and information. These stakeholders are drawn from a very wide and diverse audience, which includes government and public service agencies, decision makers and policymakers, researchers and the general public.

Several projects have provided substantial contributions on how to development the quality of input data for environmental statistics (FDES 2013). However, less effort has been devoted to the question of how-to presenting climate change Data for Statistical Purposes Based on their expectation's users and perceptions.

So, this paper will discuss the new techniques in the presenting climate change data Paving the way to statistical literacy, understanding the world through environmental data.

In this work we aim to examine the relationship between presenting climate change data for Statistical Purposes and Paving the way to environmental statistical literacy in Egypt using a set of selected measurable indicators. The selected indicators are related to air, emissions, and environmental situation in the form of infographics, graphs, and illustrations that suit the audience.

Also, the paper will discuss the results of measuring satisfaction of the audience about using these techniques in the presenting climate change Statistics data that used in annual report on environmental statistics for the year 2021 of Egypt. Moreover, an exploratory field study was conducted in December 2021.

The study deduced great variability in presenting climate change Statistics data for Statistical Purposes and Paving the way to environmental statistical literacy for general public.

Keywords: stakeholder's expectations; Information; statistical literacy; statistical quality.

Title of abstract: Everyone counts: Statistics Portugal stands for statistical literacy

José Alberto Pinto Martins, Margarida Rosa, Paula Nogueira – Statistics Portugal, Lisbon, Portugal

Statistical literacy growth is part of Statistics Portugal vision and is present in its activity in different ways. We are aware that it is not enough to produce statistical information and disseminate it effectively. It is also necessary to contribute to its understanding and wider use, aiming a more enlightened citizenship.

Statistics Portugal website is the main dissemination channel and is regularly improved in order to meet citizens and society needs, follow trends of usability, ease access and understanding of data and metadata, either using desktop PCs or mobile devices.

In order to promote statistical literacy, Statistics Portugal carries out the Open Doors Seminars project since 2015. Seminars are free and site oriented, with different themes, being targeted to general public. In 2018 this project gained an upgrading in dissemination, communication, and quality evaluation. Due to the pandemic situation this project was in standby since 2019, but we intend to maintain it in the coming years, indoors at our premises, and on- line.

Statistics Portugal develops actions for other target groups, by cooperating with educational communities in didactic projects: Local Applied Statistical Action (ALEA) and Explorística are two projects providing resources and tools that give support to the teaching/learning process of Statistics (in basic and secondary education). Both show the importance of statistics in various society aspects. The projects received the Project Award in Statistical Literacy (ISLP), in 2015 and 2017.

The main channel of ALEA is a website that offers contents and tools of scientific, pedagogical and playful interest, while Explorística develops its purpose by providing interactive games either in physical modules that can be used in itinerant exhibitions (such as in schools), or used online through a web browser, or even through apps for mobile devices.

With an identical purpose and target, but in a European level, we participate in the European Statistics Competition (ESC). Statistics Portugal is organizing for the fifth time this competition. In 2019 (ESC 2020) Statistics Portugal has invited the Central Bank of Portugal to join that organization. This event is being promoted through various channels, and a large participation from schools is expected. The 2018/2019 edition of ESC involved more than a thousand students, having decreased somewhat in the following edition, given the pandemic context.

The academic community is another target group to which Statistics Portugal has a special concern, since it has specific needs of statistical information.

Statistics Portugal established a collaboration network with Higher Education Institutions and promotes training actions periodically, physical and online, besides other initiatives.

For Statistics Portugal everyone counts – paving the way to statistical literacy is the right track to reach all users.

Keywords: educational communities; Statistics Portugal website; Open Doors Seminars; scientific community.

Title of abstract: Understanding the territory through data, data journalism and civic engagement: the Asoc project from Italy to Europe

Patrizia Collesi – Italian National Institute of Statistics, Rome, Italy
Simona De Luca – Evaluation and Analysis Unit (NUVAP) of the Department for Cohesion Policies of the Italian Presidency of the Council of Ministers, Rome, Italy

The paper presents the experience of the At the School of OpenCohesion (ASOC) project, which is an innovative educational programme aimed at promoting and developing principles of active citizenship in secondary schools.

The text will first focus on the Italian experience, with an in-depth contribution on the collaboration of Asoc with the Italian Institute of statistics, and then move on to the European experience, which started as a pilot phase in some countries in the 2019–2020 school year.

The Asoc project is an educational programme aimed at Italian schools with Open Data on projects funded by European and Italian resources, by way of projects of civic monitoring and research of European and Italian public funding for cohesion policies".

It is intended for both secondary school students and teachers. For students it provides the opportunity to develop skills in digital technologies, statistics and civics in order to help them understand and describe how cohesion polices impact the communities in which they live. For teachers it is an opportunity for upskilling in several topics, as an official recognized course for continuous learning.

ASOC has a working agreement with Istat, to foster the knowledge of the official statistics data and indicators to "measure" the projects they are monitoring. The agreement has at its core activities for both teachers and students.

On the teachers' side the agreement provides cooperation with Istat's experts in learning modules to provide basic statistical literacy competencies, on official statistics topics such as the data production process. The collaboration for students is about training on techniques of quantitative and qualitative research, on how to interpret statistical data and to build synthetic indicators and orientation on how and where to find official statistical data suitable for the monitored projects.

Some data

ASOC was launched in 2013 as part of the OpenCoesione initiative, coordinated by the Evaluation and Analysis Unit (NUVAP) of the Department for Cohesion Policies of the Italian Presidency of the Council of Ministers in collaboration with the Italian Ministry of Education and Research and the Representatives of the European Commission in Italy.

The ASOC pilot project was carried on during the 2013–2014 school year. Until now, more than 20,000 students and 1,500 teachers have been involved. The project has been acclaimed by the Open Government Partnership as a best practice to engage citizens in the policy process.

The 2019–2020 school year features the first international ASOC pilot with the support of the European Commission.

Keywords: civic monitoring; secondary-school students; teachers upskilling in statistical literacy; cohesion funds projects; statistical literacy in territorial projects.

Session 31
DATA ETHICS AND ENHANCING TRUST

Title of abstract: Quality is a Safeguard, Not a Foundation: Quality in Official Statistics and Recent Advances in Trust Theory

Ferenc Mújdricza – Hungarian Central Statistical Office, Budapest, Hungary

Trust is undeniably 'the cornerstone for any statistical system'. Namely, being trusted is essential to fulfil the role of the main provider of facts on crucial aspects of society, economy, and environment for stakeholders (citizens, academics, policy-makers, enterprises, etc). An emerging trend in academic literature, as well as national and international official statistics policies, proposes the key to building *trust* in official statistics is to maintain and enhance *quality*, among others: transparency, communication, independence, and privacy guarantees. Quality is of paramount importance and a determinant of the value of statistics. But how does it add to trust towards statistics – is it a suitable foundation for building trust in official statistics?

The proposition of quality as a 'cornerstone' of trust is in line with the cognitive approach to trust concepts. These describe trust as a knowledge-, interest- or experience-based, learnt capacity, a rational decision or expectation assuming uncertainty, risk, and vulnerability, which is hard to build but easy to destroy. Yet this approach describes a calculative, shrewd, and potentially manipulative frame of mind – the opposite of trust. These concepts mistake trust for naïveté or reliance. Creating a naïve audience risks being manipulative, which is not the aim of official statistics. Reliance should not be conflated with trust either; it only conserves a lack of trust in the worst of cases. So, trust cannot be 'built' on the bases of quality or of any cognitive factor.

Reliance, however, has a secondary function in supporting and preserving actual trust, i.e. trust as described by the non-cognitive approaches. This trust is 'a pattern in the weave of life', an innate faculty, a mostly unconscious sense of security, a tacit expectation of goodness. It is robust and highly resistant to facts and counterevidence, and usually reaches awareness when it is already harmed. Ergo, although quality cannot build trust, the lack of it may well destroy it. Quality is not a prerequisite or a basis, but still an important safeguard and support for trust.

Regarding the 'enhancement' of trust in official statistics, non-cognitive trust literature does not offer a solution. Therefore, the paper will present a novel approach to trust based on the relationship between existential anxiety and trust. The only non-pathological means of enhancing trust is to demonstrate *our* genuine trust to activate the 'virtuous circle' of trust. The relationship thus remains non-intrusive and trust-assuming, empowers the user and facilitates their relaxing into trust. Our demonstration of trust leaves the users' autonomy unharmed, so it cannot build trust in the strict sense, but it can increase the chance of them turning to official statistics with trust. The role of quality is to guard and support trust, which approach enables the development of a trust management framework for official statistics.

Keywords: official statistics, trust theory, quality, autonomy.

Title of abstract: What does statistics serving the public good mean? Evidence from the UK's statistics regulator

Dr Mary Louise Cowan – Office for Statistics Regulation, London, UK

The Office for Statistics Regulation (OSR) is responsible for regulating official statistics in the UK. OSR have regulatory tools such as assessments, systemic reviews, and compliance checks, which evaluate official statistics against the Code of Practice for Statistics. The Code of Practice contains three pillars: trustworthiness, quality, and value, and is underpinned by OSR's vision that statistics will serve the public good.

To OSR, statistics serving the public good means maximising the benefits associated with statistics. Statistics should serve the public good because they are a public asset and uphold democracy in our society. This also means that statistics should achieve the widest usage possible, so that everyone in society can benefit from them.

OSR have created a Research Programme with the aim of developing the understanding of what it means for statistics to serve the public good in practice. In the last two years, the Research Programme has begun work to understand public good from the perspective of the literature, researchers, statistics regulators, and members of the general public.

Early findings demonstrate the tensions involved in ensuring statistics serve the public good. There are different perspectives on who 'the public' is. Our evidence demonstrates researchers primarily view serving the public good as providing evidence for policy makers. This approach contrasts starkly with others who advocate for the general public and want statistics to contribute towards the public's general understanding of the world around them. In summary, we have identified different approaches to the public good, which each conceptualise the term in their own way; and which reveal different perspectives on what, and who, statistics are for.

The findings of the Research Programme will help to develop a wider understanding and more complete definition of what statistics serving the public good means. This is useful because it informs OSR's work as statistics regulators. Additionally, the statistical system may benefit from a refined understanding of what this means so they can understand how to maximise the benefits of statistics.

Keywords: the public good, regulation, research, general public, public interest.

Title of abstract: Trusted Smart Statistics and the need for new ethical principles

Yolanda Gómez, Ana Cánovas, Ana Carmen Saura – INE, Madrid, Spain

The changing environment in a datafied society pushes the statistical world into a long-distance race where the final line is never reached because the path is continuously moving along the way. Many countries all over the world are searching for new approaches, new tools, skills and new possible roles for the National Statistical Institutes (NSIs). But some horizontal issues are essential to address properly these changes, the first one is the access to Big Data (including Internet of Things – IoT) and the legislative framework and ethical principles related to such access. The second one is to communicate to the citizens these principles and to inform about the statistical treatment of the data from these new sources. The pandemic situation has increased even more the use of these data and the open questions on their uses.

Some members of the European Statistical System (ESS) have already elaborated different ethical codes and even ethical assessment tools, which could be overlapped or/and complemented the European Statistics Code of practice (CoP). In parallel under different projects and groups in the ESS and in the European Commission (related with the EU Data Strategy) some principles have been also proposed and also good recommendations and some falls under the ethical behaviour. The European Statistics Code of Practice has just been updated in 2017, however, this changing environment is demanding new rules and principles that could be incorporated in the Code or even in an amended Regulation 223/2009 on European Statistics.

In this paper we analyse the existing ethical principles for Big Data uses (in a broad sense) and in different scenarios and compares them with the current statistical principles, we will try to go even further and think in the next future, about what kind of principles we would need if we add new roles to the NSIs.

Keywords: Smart Statistics, Big Data, Code of Practice, institutional environment, legal and ethical issues.

Title of abstract: Conceptualization, design, and implementation of a statistical ethics system for official statistics of Colombia

Wilson Herrera Romero, David Hernández Zambrano, Elizabeth Moreno Barbosa, Ruth Baquero Quevedo, Erik Arciniegas Rincón – Departamento Administrativo Nacional de Estadística (DANE), Bogotá, Colombia

The generalized phenomenon in western democracies regarding the important role that statistical information has acquired in democratic debates and in the design, implementation and evaluation of public policies, leads to rethinking the importance incorporating ethics principles into the production of statistical information, those principles involve autonomy and freedom of citizens, equity and solidarity, transparency, trust, defense of the common good and security in the use of data.

So, the NSO of Colombia (DANE for its acronyms in Spanish) in its work of providing information for everyone, launched the Statistical Ethics System (SETE), an initiative that is established as a set of principles, procedures and instances that studies the ethical considerations that may exist in the phases of statistical production (according to the GSBPM), and that allows taking preventive or corrective measures to ensure an adequate management of ethical risks. This system is pioneer in the region and is inspired and based in international references, especially the experiences of the United Kingdom and New Zealand that have data ethics advisory committees.

The document is organized in five parts. The first presents the conceptualization and design phase; the second part presents the statistical ethical framework made up of four axes, fifteen principles and the identification of new ethical risks; the third part presents the structure of the System, made up of three instances, two of them deliberative; the third part presents the stages or flow of the ethical evaluation of statistical operations and the ethical self-diagnosis tool; The fourth part includes some reflections on the implementation of the system and the challenges of SETE for its consolidation, appropriation and expansion to the National Statistical System of Colombia; and the fifth part presents some closing conclusions.

Keywords: statistical ethics, principles, ethical risks, statistical production process, ethical framework.

Title of abstract: Promotion of quality assessments and their results, as an instrument for better transparency and trust in official statistics

Alma Hodo – GOPA Luxembourg, Luxembourg

Continuous interaction with users is one of the Quality framework-complementing building blocks of the European Statistical System (ESS) Code of Practice (CoP) and the Quality Assurance Framework of the European Statistical System (ESS QAF). The aim is to increase transparency within the European Statistical System by accepting and implementing these principles, indicators, methods and best practices. The responsibility and aspects used to measure quality are explained in the CoP principle 4. (Commitment to Quality). It is through assessments like the peer reviews that statistical authorities measure the extent to which the Code of Practice has been applied at institutional level, but also at specific statistical process and output level through e.g. the internal quality reviews.

Hence, continuous quality assessments (at institutional, process, and output level) and the presentation of findings and results increase the transparency and trust in official statistics. The best practices and methods to implement the quality assessments are a combination of self-reviewing procedures and the involvement of external experts, for getting an independent and objective opinion (see CoP indicator 4.4).

Information on procedures, processes, tools and methods, but also identified areas for improvement and the means used for their implementation should be made available to the public, as much as possible. Continuous increase of transparency on quality assessments will increase the trust in official statistics. Communicating to the outside the results and findings of identified areas for improvement and presenting best practices by making the maximum of information available to users, could not only contribute to the increase of trust, but also further promote the use of official statistics by the general public.

Nevertheless, it is important to bear in mind that the CoP principle 5. (Confidentiality and data protection) must be respected at all times when presenting and sharing statistical information.

This paper will present the importance of increasing transparency of quality assessments, their findings and best practices, combined with a self-critical approach aimed at building trust of users in official statistics. Users must be surveyed for collecting information on their needs in order to produce more relevant statistics, but at the same time, users must be regularly informed on how quality of statistics is measured and which improvements are regularly undertaken by the statistical authorities.

Presenting best practices on how certain procedures are implemented in different statistical aspects is a positive and quite popular approach. However, would the idea of presenting the identified areas for improvement in a self-critical and honest manner be an instrument to build even more trust? Could statistical authorities

reflect more on the possibilities to keep a regularly updated public record (e.g. inventory, library...) of identified areas for improvement at institutional level but also at the level of specific statistical processes? What is the impact on the overall quality if such information is shared with users, and to which extent could a statistical authority share such information? Could users contribute to the improvement of identified weaknesses?

This paper will contain some new challenges and possibilities on increasing transparency which statistical authorities could take into consideration.

The continuous trade off between the increase of transparency through sharing information on quality assessments, findings, best practices, and at the same time also respecting confidentiality and data protection, is a challenging task for statistical authorities. Nowadays, for the general public, statistics should not only be available, but also easy to understand and presented in an interesting and user-friendly manner. In parallel with this mission, a potentially more open and self-critical approach of the quality assessment, by involving even more external experts could result in a more objective analysis and potentially provide additional inventive approaches.

Keywords: Transparency of findings; Identified best practices; Use of official statistics; Public Inventory of improvements; Trusted statistics.

Title of abstract: Making the point for statistical competitions in Italy: some evidences about users' participation and evidences from data

Patrizia Collesi, Daria Squillante – Italian National Institute of Statistics, Rome, Italy

Making official statistics data known and understood as largely as possible is one of the main objectives of Istat's policy in communication and dissemination. Engaging the general public and the young generations is no doubt one of the best ways to reach this objective.

A wide range of competitions are available for this purpose and for different school levels.

Bringing students closer to the subjects in a playful way, also with gamification activities outside the school channel, is one of the axes along which Istat's policy has been moving.

In the guidelines/intentions of the institutional statistical literacy policy, Istat expert, the statistician who knows statistical topics, does not replace the teacher, but is an expert for those subjects that a teacher is not obliged to know and that can enrich the students' curricula and broaden their competences.

A relevant help in this sense, which promotes the knowledge of official statistics, derives from the institution of civic education as a cross-curricular topic in which elements of knowledge of official statistics can be included, as a way to expand our knowledge of the world around us through objective measurements.

To introduce official statistics, competitions and talents are proposed, with different purposes and targets, for various school levels and with different levels of difficulty. For the non-experts, for the gifted and for those who simply want to starts data knowledge.

This document presents the various types of competitions and contests for students, with numerical evidence about participation and results for well established and traditional competitions, and presents the project for the new talent on Infographics "Create your own infographics", launched as a pilot competition at the 14th National Statistics Conference.

A special focus is dedicated to the Statistical Olympiads, presenting developments and data from the last five editions, with the solutions adopted with remote testing in the pandemic period as well.

For the new talent on infographics, the project is presented, along with its objectives, the numerical evidence deriving from the first experiment and the perspectives for

evolution. The intention is to turn the activity into a stable project having the dual function of fostering knowledge of statistical data and knowing how to combine them with images.

Keywords: unconventional ways to engage new users; infographics; statistics competitions; statistical thinking; reasoning with data.

Title of abstract: Reaching a Wider Audience Through GIS

Linda Peters, Kate Hess – Esri, Redlands, CA, USA

For decades, users have employed Geographic Information Systems (GIS) as a system of record. In today's world, however, that is not enough. To make a system of record more valuable, the information it holds must be shared with a broader audience in a collaborative way. In other words, a system of record needs a system of engagement.

A system of engagement is critical today to help National Statistical Offices (NSOs) serve society and data users with the statistics they need, to engage users, and to reach a wider audience. Data users today expect to engage in ways that are convenient or natural to their workflows. Whether that is a statistician looking to download a tabular set of data, a teacher – looking to educate students with a simple map or a program developer wishing to connect to an Application Programming Interface (API) allowing them to extract the data they need for their app. Users expect to be able to gain access to accurate data through a simple easy to use interface, in a timely fashion.

NSOs have been using systems of record for decades... the time is now for systems of engagement offering us new ways of statistical communication. Successful systems encourage peer-to-peer interactions, and leverage cloud and mobile technologies wherever possible to enable interaction in a scalable and cost-effective way.

The ArcGIS system of record allows you to provide complete, accurate, timely and consistent data. Coupled with a system of engagement, it helps us facilitate and orchestrate the data user journey. Esri's System of Engagement is designed so that everyone can access information in the way they need to, whether that is accessing small area data, a map, a chart or other type of information graphic or an API. GIS systems provide tools for data users to answer questions and apply statistics in decision and policy making more effectively.

Systems of engagement include the programs and applications that workers interact with in their daily work lives. We will in this session look at several key GIS applications such as Hub and StoryMaps that leverage the power of geography and location integrated with statistics. The power of geography is that nearly anyone can look at a map and quickly grasp complex information. By leveraging ready-to-use apps, content, capabilities and GIS infrastructure, everyone can visualize statistical data as well as analyze, and collaborate using maps and information graphics in a GIS system.

Keywords: Trust, Authoritative, Data Users, Engagement, GIS.

Title of abstract: Disseminate more with less: an utopia? or the ambitious aim of Insee

Alexandre Duval, Frédéric Minodier, Christine Lagarenne – INSEE, Montrouge, France

INSEE is conducting a project aiming at redefining the whole process of data dissemination. Actually, disseminating data requires a high quality of metadata, regarding accuracy, timeliness and so on. This is also required for international comparisons. The objectives of the project are both internal and external, trying to erase boundaries generated by separated production processes to offer a unique point to get all the Insee's data.

The inside part is to improve the efficiency and the reliability of the dissemination process by putting metadata at the core of it. First attempts are made upon yearly national accounts, population census, tourism: major findings will be exposed, regarding the results and the way to reach them. The external part aims at offering a central point on our website from which users can discover all the figures and navigate through them. Based on international standards like Stat-DCAT-AP and deeply linked to metadata, the project will offer a range of services starting with a datacatalog and going to dynamic datavisualisations. A first realization will be shown.

Keywords: metadata driven dissemination, datavisualisation, official statistics.

Title of abstract: Dashboard in information dissemination practice: the Weekly Monitor

Krisztián Kovács – Hungarian Central Statistical Office, Hungary

The pandemic has increased the demand from economic actors and data users for quick, simple and concise data releases. In response to the changing user needs, the HCSO has reacted rapidly by expanding its product portfolio with new elements.

Following the emergence of the COVID19 pandemic in Hungary, in May 2020 we created a sub-page on the website called "Coronavirus Dossier", where we collected statistical information on the impact of the coronavirus pandemic. The dossier includes analyses, data visualisations and information on the H-UNCOVER representative survey.

We complemented our analysis with estimates of the impact of the Covid19 outbreak. To do this, we used an outlier estimation that can be easily and quickly calculated for any time series for which we apply a seasonal adjustment.

During this pandemic, timeliness has become more important than ever. Not only did we speed up processes, but we also broadened data reporting. In the wake of the outbreak of the coronavirus pandemic, we found it necessary to offer more frequent and up-to-date statistics in a number of statistical areas, in addition to the regularly published monthly or quarterly economic indicators.

To achieve this, in May 2020, we were among the first EU countries to include dashboards in our product range. One such publication is the HCSO Weekly Monitor, a simple HTML-based dashboard the value of which is not in the technology but in the content. In the first year we focused on the pandemic and its socio-economic impact. Today, it has been transformed into a dashboard that shows users the current economic trends.

The interactive and downloadable graphs and maps in the dashboard are mainly created with Highcharts/Highmaps, so they are relatively easy to produce. We have created a CMS (Content Management System) system where editors can produce typical graphs without deep programming knowledge. Due to the scarce IT resources, we chose a software that does not require any maintenance and that uses a version control (git) to save the content. This ensures the traceability of the content.

The dashboard has a high download rate and is often referred to in the media. The most visited topics included Covid survey data, Prices and Demographics.

Overall, the dashboard-style presentation of data can be considered a good practice in international comparison, because there is a strong demand from users for a meaningful presentation of data. This is enhanced by headlines that highlight and draw attention to the current trend or situation.

Keywords: dashboard; information dissemination; pandemic; interactivity; timeliness.

Title of abstract: Statistical data – here and now during the times of Covid or other crisis

Laima Grižaitė – Statistics Lithuania, Vilnius, Lithuania

At the end of 2020, Statistics Lithuania took over calculation and publication of COVID-19 indicators. Statistics Lithuania is most likely the only NSI in the World, which took leadership to manage and publish operational COVID-19 data. This courage paid off: we got new users of official statistics, twice more followers in social media, duplicate numbers of quotations in traditional media.

For the publication of operational pandemic statistics we used interactive dashboards.

The ESRI Experience Builder tool was used to create an application for public users, which provides a set of dashboards on COVID-19 statistics. The application is also adapted for mobile devices. It is widely used and is viewed tens of thousands of times daily.

The dashboards can be used to monitor not only cases, tests and deaths but also the course of vaccination, i.e. the proportion of the vaccinated population and distribution thereof by age group and municipality. Municipal administrations also monitor the information on vaccination doses: the number of vaccines received, distribution thereof by municipality. At the request of users, the dashboards also provide information on the use of vaccines by manufacturer.

The published information on outbreaks has also received much attention: users can monitor where and in which establishments there are new outbreaks of COVID-19 and where the outbreak is already subdued.

The ESRI platform is also used for publication of timely open data, which are growing in popularity every day. Open data provide more detailed information on COVID-19 cases, testing and vaccination, as well as data to predict emerging outbreaks.

Such an understandable, easily accessible and open presentation of data builds public confidence and facilitates the understanding of state decisions on quarantine conditions.

This solution allowed Statistics Lithuania to become more visible. Leadership in data management has strengthened the brand of the institution, especially among decision makers and analysts, to name the paths in the production of operational statistics.

Keywords: Dashboard, ESRI platform, publication of data, leadership in data management.

Session 33
PRODUCING NEW DATA IN COVID TIMES

Title of abstract: NSI in the provision of critical data in the COVID-19 crisis: The use case of vaccination statistics in Austria

Valerie Holzer, Franz Astleithner, Johanna Einfalt, Dr. Regina Fuchs, Christine Ning, Manuel Reif, Guido Sommer-Binder – Statistics Austria, Vienna, Austria

The COVID-19 crisis has illustrated the importance of timely and detailed data to inform society and policy makers about important topics. National Statistical Institutes (NSI) can play a key role in the provision of these information. They are equipped with expertise in various fields that are important for the provision of relevant data during the crisis, comprising experience in linking of various data sources and meeting high data privacy standards. We will present use-cases of vaccination statistics that were carried out by Statistics Austria in 2021. In this paper, we focus on the use-case of students and teachers in educational institutions. Vaccination data is of high importance, when it comes to evaluate closure and opening of schools.

We will explain in detail the methods used and some of the challenges and lessons learned in the process of creating vaccination statistics. Such as how to deal with source data sets with different reference dates, defining the scope of the published results and find appropriate methods to guarantee data privacy.

After completion of a project, not only the client but also the population were informed about the results. Moreover, our use cases illustrate that in order to have access to the necessary data, new laws had to be created at the beginning of the COVID-19 crisis. Due to the urgency of the topic and due to the high level of trust in the NSI the necessary laws were passed comparatively quickly.

Our case studies show, how NSI can deliver important information that in the end may help policymakers to evaluate COVID-19 based risks in schools and thus provide the opportunity to reopen schools sooner than initially planned. What we take away from our use-cases is that it is particularly important to have a legal basis to be able to quickly respond to potential future crisis.

Keywords: vaccination statistics; register-based statistics; National Statistical Institutes – NSI; COVID-19 crisis.

Title of abstract: Monitoring mortality shocks across time and space

Dmitri Jdanov – Max Planck Institute for Demographic Research, Rostock, Germany
Domantas Jasilionis – Max Planck Institute for Demographic Research, Rostock, Germany,
Demographic Research Centre, Vytautas Magnus University, Kaunas, Lithuania

Effective public health responses to epidemics have always required timely and reliable monitoring of the situation. During the last decades, there have been numerous short-term mortality outbreaks related to influenza, heat waves, natural or made-man disasters. These were warning signs calling for a new approach in collecting mortality data for better understanding the nature and public health consequences of such outbreaks. Nevertheless, the COVID-19 pandemic revealed the lack of progress in establishing efficient international and national monitoring systems for providing timely and accurate data that could be used to trace the pandemic course.

The main challenges of such monitoring are related to 1) uncertainty of the key epidemic parameters (e.g. case fatality risk or cause-specific mortality ratio) and 2) ability to estimate both direct and indirect losses caused by the respective short-term risk factor. This most reliable approach to quantify the total direct and indirect mortality burden of the pandemic (as well as any other short-term risk factor) is estimating weekly excess mortality. The excess mortality is calculated as a difference between observed and expected levels. Fortunately, almost all developed countries are able to produce weekly mortality statistics with minimal delay. It took only a few months to make the weekly death counts available in the public domain. Therefore, collecting and studying reliable and timely data on short-term mortality fluctuations become an important evidence base for monitoring and addressing unexpected health threats. However, this important change in population statistics (which we consider at the next stage of the data revolution in demography) generates new challenges for science and society.

First, there is a trade-off between needs in detailed statistics to build evidence-based public health policy and personal data protection. Second, there are no standards in collecting such data, and the quality of statistics varies across countries. Third, there is no standard methodological guideline for the calculation of excess mortality. In this paper, we present the Short-Term Mortality Fluctuation (STMF) data series, an extension of the Human Mortality Database Project. The STMF is the first and (at the moment) the only scientific database that provides reliably weekly death counts and death rates by age and sex. We discuss possible data limitations, methodological challenges, and present an original methodology that we apply to get detailed and comparable data. We also examine the sensitivity of excess mortality metrics to the choice of the mortality index, method, reference period, and the time unit of the death series. Using the data from the STMF, we analyze COVID-related total losses in 38 developed countries in 2020 and 2021.

Keywords: Covid-19; weekly excess mortality, monitoring pandemic.

Title of abstract: Unconventional surveys for unprecedented times: the Eurofound Living, Working and COVID-19 e-survey

Daphne Nathalie Ahrendt, Eszter Sandor, Tadas Leoncikas, Massimiliano Mascherini – EUROFOUND, Dublin, Ireland

Collecting timely information during COVID19 was essential for policymakers to understand the economic and social dynamics of the crisis and to promptly prepare policy responses to smooth its impact. However, the implementation of lockdowns imposed to control the COVID-19 made the use of traditional data collection methods, i.e. face to face interviews, broadly impossible.

Since March 2020 a surge in the use of online surveys to collect information was recorded. Online surveys offer a cost-effective and faster way of data collection compared to other data collection methods; both cost and speed are important for getting topical results, especially in context of the crises. Alongside these advantages, there are several concerns to address regarding validity and generalizability of findings as well as regarding possible data privacy and ethical issues.

In April 2020 Eurofound launched its *Living, working and COVID-19* e-survey (LWCOVID19) to capture the impact of the pandemic on people's lives and work in Europe. The e-survey was fielded in all EU Member States with a questionnaire translated to all main EU languages. As far as we know the LWCOVID19 was the first e-survey launched by an EU body during the pandemic providing comparable information for all the Member States.

Started as standalone exercise, due to the developments of the pandemic to date already four rounds of the LWCOVID19 survey have been fielded. Overall, more than 150,000 European respondents, recruited through social media advertisement and snowball sampling, have taken part in the surveys. A panel component of around 45,000 panel members makes it possible to track people's situation across the four rounds. A fifth round of the survey will be fielded in spring 2022.

In a broad context of social statistics, the LWCOVID19 exercise filled in a niche for urgent and topical data with efforts applied to address quality challenges. Since LWCOVID19 used a range of variables also applied in other sources, including Eurofound traditional surveys, validity concerns could be addressed. LWCOVID19 included also some EU-SILC items (such as difficulty making ends meet and financial fragility) so that early trend indications could be captured and ex-post comparisons between the data sources explored. Furthermore, it fielded new questions, such as on telework, trust or vaccine hesitancy that could eventually inform policymakers as well as the choice of new topical variables in official statistics.

This paper will focus on the design of LWCOVID19 large-scale online non-probabilistic survey, on the lessons learned along the way and on the improvements that have been made to date. In particular it will cover the various stages of the survey development and zoom into the quality of the data, the weighting methodology, the

use of social media campaign to reach participants, the panel component, and the use of reminders and incentives as well as data privacy and ethical challenges.

Keywords: online survey; COVID19; panel survey; proactive data collection; quality of data.

Title of abstract: 'But what does this mean?': Generating trust and understanding to influence policy and inform the public

Sarah Crofts, Melissa Randall – ONS, Titchfield, UK

Throughout the covid pandemic data have been plentiful, with case data, hospitalisations, deaths and a variety of vaccine effectiveness information flooding news sources. It can be hard for users and the public to understand what they mean, and to navigate and choose the correct information for the question they are trying to answer.

At the same time, it is vital that decisions are made using good quality data that suits the purpose it is put to, and that the public can access the same to hold decision makers to account.

The Office for National Statistics' COVID-19 Infection Survey (CIS) has delivered weekly, near real-time insight on infection and immunity which goes directly to the heart of Government. It has been used to make key decisions to navigate the pandemic and is regularly featured in the news to inform the public.

The CIS is the largest regular survey of coronavirus (COVID-19) infections and antibodies in the world. It covers the four countries of the UK.

The Office for National Statistics (ONS) has worked extensively to generate trust in the analysis, and communicated clearly in our publications, blogs, press releases and media interviews to ensure trustworthy, reliable statistics are available for policymakers and for the public. A recent survey has shown that 15% of the British public have visited the ONS website to look for information on the pandemic, which demonstrates an appetite for information directly from source that we haven't seen before.

This paper will share examples from the range of activities which have ensured that all users, including experts, decision makers and the public, can find the information they need, can clearly understand what our data are telling us and trust what we say. Examples will take two forms – firstly, how we have worked with our users to ensure we meet their needs. Secondly, examples of ways of communicating data to ensure it is understood and trusted.

Keywords: trust; users; quality; covid.

Session 34
CHALLENGES FOR STATISTICAL PRODUCTION IN COVID-TIMES

Title of abstract: The challenges of Statistical Business Registers during the COVID-19 pandemic

Péter Pálosi – Hungarian Central Statistical Office, Budapest, Hungary

The Statistical Business Register (SBR) is the backbone of official business statistics. A properly managed and maintained SBR is crucial in developing and compiling high quality official business statistics. An SBR is a system for supporting public administration and statistical data collection which also handles the various needs of business statistics. The SBR has various data sources, and it is being updated with administrative data from different organisations (e.g. Registry Court, Tax Office) daily. Due to the COVID-19 pandemic, the importance of administrative data, including the role of SBRs significantly increased, because the administrative data is fast, timely and it is not a burden for businesses. The daily data of new business registrations and bankruptcies are turned out to be a major input for policy makers and for the Statistical Recovery Dashboard. The increased interest in the SBR has brought to the surface the challenges posed by the complexity of the system. The main challenge of developing and maintaining an SBR is in its sheer complexity. Its core logic is different in each country due to the national legal systems. The development and maintenance of an SBR requires knowledge of different fields such as economics, statistics, law and IT. A small change in a legislation governing business units may greatly impact the logic of the SBR, therefore a strong cooperation not only between the departments within the statistical office, but also with the partner organisations is vital. Organisational solutions respond differently to the rapid changes required by the environment.

During the COVID-19 pandemic, the Hungarian government issued different legislative changes aiming to support the activity of private entrepreneurs. A recent change allows private entrepreneurs to select a provisional date when they wish to stop or suspend their activities. The SBR methodology doesn't allow using provisional dates, therefore the SBR's inner logic had to be changed. In the Hungarian experience, the biggest challenge is coordinating the work of colleagues who are responsible for the SBR, but working in different departments, mostly IT experts and statisticians. There has to be a strong bridge between the IT department and the Register Section of the Statistics Directorate responsible for SBR methodology: the Register Section needs to have a fundamental IT and programming knowledge as well as the IT department has to be familiar with SBR methodology to be able to develop and maintain the SBR. In conclusion, the changes caused by the COVID-19 pandemic simultaneously focused the attention on the benefits of the SBR's timeliness, and on the sensitivity of the system through legislative changes, which put the role and position of the SBR in the public administration and in the statistical offices again on the agenda.

Keywords: Statistical Business Register; COVID-19; administrative data; official business statistics; organizational structure.

Title of abstract: Governments measure during the Covid crisis and statistical implications in national accounts: the case of tax deferrals

Susanna Riccioni, Luisa Sciandra – ISTAT, Rome, Italy

During the COVID-19 crisis many European governments have implemented major policy measures to support companies and households coping with the sharp decline in economic activity. Correctly interpreting and recording these measures in the national accounts was a major challenge for compilers who had to ensure not only timely but also independent, reliable, and comparable statistics in exceptional circumstances. High quality official statistics are fundamental for public finance monitoring, both at quarterly and at annual level. They provide informative basis for policy makers and economic analysts and play a central role in the context of the European fiscal surveillance process. This paper discusses how the Italian Statistical institute (ISTAT) managed to quickly respond to the sudden increase in the need for public finance information in the aftermath of government interventions to contrast the consequences of COVID-19 crisis.

In particular, the paper focuses on one of the first fiscal actions Italian government enforced immediately following the outbreak, that is deferrals of tax obligations. According to OECD (2021) this is the main fiscal relief European countries have introduced in response to Covid crisis. The urgency of giving appropriate trace of these actions in the framework of national accounts has raised a significant methodological issue for compilers, and its solution can be exemplary of an innovative approach to provide data during extraordinary periods remaining as compliant as possible with the conventional processes and codified rules. The new approach is based on the combination of new data sources with already existing data, on a strengthen interaction among data providers and government agencies to assure a speedier access to administrative data (Ministry of Finance and Revenue Agency), and on a strict collaboration with European institutions (Eurostat).

The paper also discusses some critical aspects emerged during the crisis and underlining the relevance of timely and accurate statistics.

Keywords: Covid-19, taxes, general government accounts.

Title of abstract: Consequences of the COVID 19 pandemic on Household Surveys Data Collection

Víctor Balsa Criado – National Statistics Institute of Spain

The COVID pandemic appeared unexpectedly in March 2020 with many of the main household surveys collected largely using the face to face interview. This paper details the actions that (due to confinement and the impossibility of conducting face-to-face interviews) had to be improvised to convert the CAPI-CATI model into a mixed CATI-CAWI model quickly, efficiently and fulfilling the survey deadlines.

The response rate, and above all, the rate of surveyed dwellings over gross sample, logically, have been affected by this drastic change; however, some measures have clearly mitigated these effects.

Some of these measures as changes on the sample frame or implementing of CAWI interview, were already programmed for the near future, but events took over and that had to be applied on the fly.

Some other actions had to be improvised as the improvement in the processes of obtaining telephone numbers for the households on the first interview.

Another important collateral effect of the pandemic has been teleworking, either on interviewers as in interviewees.

The effects that teleworking has had on interviewers and interviews are described.

There have been observed changes in some of the main figures to measure the effectiveness and performance of the CATI: as number of calls per interviewer/day, number of households resolved by interviewer/day, daily time "on call", and others

As a result of this experience, the mixed model for important surveys such as the Labour Force Survey has definitively changed, establishing as definitive some measures taken of a temporary nature.

Keywords: pandemic; teleworking; frame; data collection mixed model

Title of abstract: Could timeliness be improved by using an online access panel? Case study of the pandemic fatigue survey

Nejc Berzelak, Maruša Rehberger, Darja Lavtar, Metka Zaletel – National Institute of Public Health, Ljubljana, Slovenia

Accurate representation of a target population is an important criterion of survey data quality. While probability samples provide the statistical basis for inference, surveys based on them face increasing challenges of declining response rates, rising costs and long data collection procedures despite the growing use of web surveys. These issues were further aggravated by the COVID-19 pandemic that limited personal access to respondents and made rapidly available data crucial for governmental and public health response. This paper addresses the use of online access panels in official statistics when surveys on probability samples are not feasible due to demands for timely provision of data and limited resources. Such panels are usually recruited with non-probability sampling and their members regularly participate in web surveys about various topics. While this enables fast data collection at significantly lower costs, it raises concerns about the coverage of certain groups of population and data quality.

National Institute of Public Health, having the role of Other National Authority for health and health care statistics in the Slovenian statistical system, conducted a study SI-PANDA to continuously monitor the pandemic fatigue of the Slovenian population. Nineteen survey waves were performed in a year. Prompt development of questionnaires was needed to respond to rapid changes in COVID-19 measures and researchers were urged to provide data from each wave as quickly as possible. To meet the required timeliness and available budget, a commercial online access panel was used. Around 1,000 panellists completed a web questionnaire per wave. In parallel, a mixed-mode survey with a comparable questionnaire was performed on a probability sample of the general population.

The paper compares the results from the probability survey (n = 3,276) and five panel waves performed at the similar time. Several items, composite measures and statistical parameters are compared to identify differences between the two survey approaches. The results reveal substantial variation of differences in estimates among variables. While most estimates are rather comparable, a minority of items exhibit profound differences that may significantly alter conclusions drawn from the results. Observation of response behaviour indicators (response times, extreme and mid-point responding and straightlining) reveal relatively low incidence of potential response quality problems, although respondents to the probability survey tended to choose higher extreme scale points more frequently. Potential reasons for the observed differences and implications for research practice in official statistics are discussed.

Keywords: online access panels; web surveys; survey quality; timeliness; COVID-19.

Title of abstract: Risk assessment of adverse drug reactions to improve quality of adverse events reporting in a large university hospital

Dovilé Ramanauskaitė, Greta Burneikaitė, Laimis Dambrauskas, Edita Kazėnaitė – Vilnius University Hospital Santaros Klinikos, Faculty of Medicine, Vilnius University, Vilnius, Lithuania

Roma Puronaitė – Vilnius University Hospital Santaros Klinikos, Faculty of Mathematics and Informatics, Institute of Data Science and Digital Technologies, Vilnius University, Vilnius, Lithuania

Justas Trinkūnas – Vilnius University Hospital Santaros Klinikos, Department of Information Systems, Vilnius Gediminas Technical University, Vilnius, Lithuania

Rimgaudas Katkus – Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania

Audronė Jakaitienė – Faculty of Mathematics and Informatics, Institute of Data Science and Digital Technologies, Vilnius University, Faculty of Medicine, Vilnius University, Vilnius, Lithuania

The State Medicines Control Agency is collecting data on adverse drug reactions (ADRs), which is the one of major pharmacovigilance components. Different ADR assessment methods have been reported in the literature, however there is no universally agreed tool for ADRs analysis. Also, it is not clear what exact data should be collected to estimate ADR. It particularly reflects the fact that not all reactions are reported completely and in a timely manner. In order to report ADRs in clinical practice and to ensure sufficient reporting, we initiated a biomedical study to disclose the characteristics of adverse drug reactions: frequency, type, predictability, severity and avoidability, impact on patient morbidity, and burden on the patient or health care system. Alongside with this task we try to evaluate the applicability and suitability of available risk assessment scales for day-to-day practice. We have included additional information in the automated data submission form integrated into the hospital information system to the State Medicines Control Agency, which allows for a detailed assessment of the risk of adverse reactions and the use of machine learning algorithms for automated risk assessment in the future.

Keywords: adverse drug reactions; risk assessment; reporting quality; pharmacovigilance.

Title of abstract: Temporal disaggregation in COVID time: a smart extension of Chow-Lin method for quarterly national accounts purposes

Rafael López Fernández, Andrés García Carreira, María Ferrero Soler – INE-Spain, Madrid, Spain

The impact of the COVID-19 pandemic is present in our societies since the first quarter of 2020. It has caused an overturn to a number of socioeconomic variables and has posed countless challenges, hitherto never seen before, in the statistical production quality.

From a macroeconomic point of view, users look in the national accounts for the picture of this new, unexpected and genuine economic crisis, so they are, once again, specifically demanded. And within this framework, the quarterly national accounts are the instrument called to give the earliest answer on the short-term evolution of the national and world economies: the shape of a crashing fall of the national and world economies and the subsequent recovery.

Quarterly national accounts rely heavily on mathematical and econometric methods. Temporal disaggregation and extrapolations procedures are in the heart of their compilation process. And these methods are based on the stability and robustness of the correlated patterns of statistical series, which have been completely altered through the successive waves of the pandemic.

INE-Spain proposes a new extension of the classic Chow-Lin temporal disaggregation method (Chow and Lin, 1971) that avoids undesirable distortions caused by the model on the preceding accounting periods due to the impact of the extreme volatility on 2020 statistical series. This innovation has been essential to keep the required quality on the recent updating of the 2018–2020 Spanish national accounts quarterly series, that provides a first macroeconomic insight on the sudden appearance of the virus in Spain.

Keywords: COVID-19, quarterly national accounts, temporal disaggregation, Chow-Lin.

Title of abstract: The use of short-term business statistics (STS) for the monitoring of industries intensive in the use of Intellectual Property Rights

Carolina Arias Burgos – European Union Intellectual Property Office (EUIPO), Alicante, Spain

The EUIPO (European Union Intellectual Property Office) and the EPO (European Patent Office) have carried out several EU-wide studies on the contribution of IPRs to the EU economy. These studies assessed the combined contribution of industries that make intensive use of the various types of Intellectual Property Rights (IPR) based on structural business statistics (SBS) at its most detailed level (NACE classes) and are updated every three years.

In general, the structures of big economies such as the EU do not change dramatically in the short term, but the recent economic crisis affected every region in the world as a consequence of the COVID-19 pandemic, and this has completely changed the picture. Therefore, in 2020 the EUIPO started to monitor the economic performance of IPR-intensive industries with monthly indicators based on short-term business statistics (STS) with a 2-month delay. These IPR indicators are updated quarterly to monitor the recovery phase during 2021 and beyond and compared with the overall GDP trend in the EU and the four largest member states. IPR-intensive industries are procyclical with higher fluctuations in recessions and also during recovery phases.

The different composition of each IPR indicator is reflected in its trend, with patent- and design-intensive industries more affected by the sharp decline in 2020 but also faster recovery in manufacturing industries in the first half of 2021. A slowing down of its annual rates in 2021 could be partially explained by the global chip shortage and other supply chain problems affecting vehicle production across Europe.

On the other side, more than 90% of the copyright-intensive industries stems from the service sector indicators and it is the first indicator reaching the pre-crisis level at the end of 2020.

The behaviour of IPR-intensive industries in the four largest EU Member States shows a similar trend but with different magnitudes in their growth rates.

Additionally, the same STS indicators are used to estimate a set of diffusion indexes as leading indicators that provide early signs of the future trends of IPR-intensive industries. They do not allow the forecasting of the expected value of the IPR indicators in future months, but they show when the indicators will change from an acceleration to a deceleration phase (or vice versa) with up to 7 months of advance.

Therefore, STS indicators are used first to analyze the economic impact of the COVID-19 crisis on IPR-intensive industries based on the size of the recession and recovery phases' annual rates of change. Secondly, the same indicators are used to complement the previous analysis focusing on the timing of those phases and the number of industries trending upward in each month.

Keywords: intellectual property rights; short-term indicators; diffusion index.

Title of abstract: Data mining and exploratory analysis of the effects of COVID-19 on the use of ICT and E-commerce from Greek companies through official statistics

Stratos Moschidis, Christina Karamichalakou, Athanasios C. Thanopoulos – Hellenic Statistical Authority, Piraeus, Greece

The purpose of this paper is to investigate the effects of the COVID-19 pandemic on the use of Information and Communication Technologies and e-commerce by Greek enterprises through officially recorded data of the annual ICT survey of the Hellenic Statistical Authority (ELSTAT). The analysis was made with the use of anonymized microdata for the common surveyed enterprises of the reference years 2019 and 2020 and was based on the common characteristics/questions of the questionnaires. Data management techniques and methods of exploratory statistical analysis were applied in the study. The results highlight changes that indicate opportunities and / or constraints for Greek entrepreneurship as regards the adaptation and evolution of ICT usage during periods of crisis and abnormal operation of the market and businesses. The results can also serve as a basis for discussion on potential methodological changes in the design of the ICT survey questionnaires that could decrease the burden on respondents and increase the consistency and efficiency of the survey.

Keywords: official statistics; data mining; exploratory data analysis; ICT.

Title of abstract: Measuring labour market participation in Europe during COVID: the role of Short-Time Work schemes

Cristina Lafuente – Universidad Carlos III, Madrid, Spain and European university Institute, Florence, Italy
Astrid Ruland – European university Institute, Florence, Italy

In this presentation we investigate the impact that Short-Time Work (STW) schemes have had on the classification of working age adults as out of the labour force (OLF) in European labour force surveys during the COVID emergency of 2020. Most European countries expanded or created new STW schemes at this time to alleviate the negative impact of economic and sanitary measures (i.e. lockdowns) from the second quarter of 2020 onwards. Applying the prevailing classification rules, we find that some workers under STW schemes who report working zero hours are classified as OLF. This is likely to be happening because some workers report that the reduction in their hours is expected to last for more than 3 months, which implies that they are not employed according to European regulation.

However, many workers in this situation subsequently return to employment (on a STW) in the next quarter. We argue that for many economic applications, these workers should have been classified as employed in the previous period. These workers have a minor impact on overall activity rates and employment, but they distort the labour market flows in and out of inactivity substantially. This is important

for measuring job destruction and job creation accurately. Classifying these workers as OLF overestimates the employment loses during COVID and the employment gains immediately after it. The same issue is present during the Great Recession, although STW schemes were more limited at that time. Furthermore, workers in industries that were more affected by STW schemes in the Great Recession are less likely to be classified as OLF in the COVID recession. We propose a simple approach to correct for this effect when working with labour market flows during COVID affected periods. For the time being, we focus on Spain and France, however we are planning to expand the analysis to Italy and the Netherlands by March next year.

Keywords: labour force survey; labour market flows; short-time work; inactivity.

Title of abstract: Early estimates of income inequalities during the COVID-19 pandemic in 2020

Leulescu Aura, Mihaela Agafiței, Adriano di Guglielmo, Andrea Gallelli – European Commission, Luxembourg, Luxembourg

The COVID crisis highlighted the need to provide relevant data in the very short term to enable policy makers to assess the effects on the income distribution in 2020. The ESS strategy for providing more timely statistics on income relies on two pillars: 1) improving the timeliness of the main data source on income, EU-SILC, and 2) providing model-based flash estimates about income 6 months after the reference period.

While model-based and still experimental, flash estimates allowed during COVID19 to provide an early assessment of the distributional impact of the crisis as well as the effects of the latest social policies put in place. Eurostat released advanced estimates 2020 on income indicators based on nowcasting techniques, which combine labour dynamics and social policies.

This paper presents the results and the improved nowcasting methodology put in place in order to capture the income evolution during the pandemic. First, we describe the labour transitions methodology used to estimate the evolution of income from work. Labour transition model-based data are produced by Eurostat, using detailed distributional information on the loss of jobs and short-term work schemes from the Labour Force Survey and administrative data. The impact across different categories of individuals, the duration of unemployment/absence and percentage of hours worked are modelled using the EU-LFS longitudinal and quarterly loops. Second, government transfers are simulated via the EUROMOD tax benefit model, which takes into account the most recent policy changes, including wage compensation measures introduced during the pandemic for preserving jobs. It is important to note, that the estimation takes into account all these components, based on an integrated methodological framework.

Keywords: flash estimates; nowcasting, income inequalities; labour dynamics; social policies.

Session 36
STAYING RELEVANT, ENHANCING RELEVANCE

Title of abstract: New quality challenges in official defence statistics: French statistical survey on « defence industries businesses »

Wyckaert Matthieu – Ministry of Armed Forces, Paris, France

In France, the Economic Observatory of Defense (OED in French) is the Ministerial Statistical Office (SSM in French) of the Ministry of the Armed Forces. It exercises its competence in the field of the armed forces and companies under national defense.

The internationalized industrial dynamic, backed by the sovereignty policies of States, has led to the emergence in the defense industry of large transnational groups in Europe. These large groups constitute with other smaller companies, including SMEs, the Defence Industrial and Technological Base (DTIB). Although the importance of the defense sector is recognized, its measurement remains yet problematic. Indeed, the nomenclature of activity in the European Community (NACE) or its French version (NAF) only partially allow us to measure the weight of the defense industry. The dual nature of the technologies used to produce war materials makes it difficult to separate their civilian, military or security uses from official statistics alone.

However, the DTIB is an object of public policies as well as subject of reporter and newspaper articles. It is an exporting industry, an employer industry and a source of major technological innovations. It is also an instrument of foreign policy or international cooperation, especially European. ESP multiannual action plan (MAP) for 2021–27 foresees to "carry out feasibility studies and based upon their assessment, develop indicators to assess the size and/or connectedness of the European Defence Industry to support policy initiatives in this area".

Access to sensitive administrative data (including customs data) and recourse to experts constitute a first approach to demarcate the productive fabric contributing to the production of war materials. However, this approach remains fragmented due to inter-company relations not subject to administrative formalities. The alternative approach is to conduct a survey. This is why the OED implemented in 2018 a first survey of official statistics on the subject. The sampling base is constituted from administrative data and the information already present in the official statistical bases. Pairing with other statistical surveys makes it possible to limit the size of the questionnaire and therefore to limit the response burden for companies.

The success of the survey launched in 2018 made it possible to estimate the importance of the DTIB in France. In 2017, the 1,190 industrial companies of the DTIB generated sales of 29.6 billion in the arms segment. More than half of military revenue comes from the sale and maintenance of aircraft or warships or their

components. Research and development for military products concerns 21% of companies. The future survey on the DTIB, called the survey on defense companies in industry and services, will include services in a way that takes into account the whole spectrum of the sector.

Keywords: Measure of French defense industries; new survey on military markets; European defence industry; French DTIB, globalization of industry.

Title of abstract: Assessing the Level of Undercounting in the International Migration Flows Reported by Eurostat

Maciej J. Dańko, Emanuele del Fava, Emilio Zagheni – Max Planck Institute for Demographic Research, Rostock, Germany

Arkadiusz Wiśniowski – Social Statistics Department, University of Manchester, Manchester, UK

Undercounting is a serious issue related to migration data quality as it may lead to bias in the migration estimates and, as such, it should be accounted for in any migration model (e.g. IMEM). It is commonly believed that undercounting is mainly associated with, but not limited to, emigration flows. In those countries using population registers or other administrative registers as main source of official migration statistics, undercounting can be thought of as the consequence of individuals failing to register when entering a country or to de-register when leaving. In other cases, when surveys are used instead as source of migration data, undercounting can result from individuals who experienced migration failing to answer the survey. In either case, all these individual decisions are likely affected by the varying requirements between and within countries for reporting of migrating individuals, or by the existing difficulties in enforcing such legal requirements.

Previous migration models assumed a time-independent undercounting country classification based entirely on expert opinion. Here we improve such classification by combining the expert opinion with both metadata and a model. The model uses the bilateral migration data to compare flows of the same direction, but reported by different countries. This paper is accompanied by a Shiny app called UndercountMigScores (https://github.com/MaciejDanko/UndercountMigScores), which provides many options for estimating and classifying the undercounting.

Keywords: undercounting; Eurostat; migration; bilateral flows, data quality.

Title of abstract: A new way of looking to territories through official statistics: place-based inequalities in Portugal

Rosário Mauritti, Maria do Carmo Botelho, Nuno Nunes, Sara Franco Silva – Iscte – University Institute of Lisbon, Portugal

The COVID-19 pandemic has exposed the complex relationships between the global, national, regional and local levels of the European social space, revealing the systemic and cumulative effects of different mechanisms of inequality on people's living conditions and well-being. In this context, the availability of reliable and timely social statistics assumes an increasing centrality for scientifically supported policy action. In the global-digital society, the mega-production of data, its open dissemination and often vulnerability to misinformation, requires up to date official statistics for robust decision-making processes acting on collective challenges.

Reversing long-standing trends of increasing inequality is one of those main challenges in our time. Inequality is at the heart of other major challenges such as climate change, global health, the digital transition and the future of work, inclusive participation and consolidation of democratic systems, all of which impact on local spaces and the people who live there. In order to understand the causes and consequences of inequality, and to put in place effective political-institutional mechanisms for their reduction, it is essential to have a multidimensional, integrated, and multi-scaling system of indicators.

As a contribution to this perspective, we present a new look at territories, considering their systemic, structural, institutional and interactive relations, which manifest in a relevant way a set of multiple inequalities. Based on official statistics from mainland Portugal, we developed a quantitative multivariate analysis, namely hierarchical and non-hierarchical cluster analysis, comparing its 272 municipalities. This place-based approach was supported by 12 inequality indicators from INE (Statistics Portugal), potentially replicable to other European countries. Official statistics were used at national, regional and municipal levels. Data from Census 2001 and 2011 and regional and municipal statistics from 2016 were used as reference, and also data from the Personal Income Tax (IRS) from 2016.

In Southern Europe and in the European context, Portugal is one of the most unequal countries. Portuguese society is crossed by five territorial configurations of inequalities, forming common features and distinctions between municipalities and regions. These territories of inequality result from complex intersections between social class, income, education, demographic density and age structure. The country is formed by «Industrial Territories in Transition», «Intermediate Territories», «Urban Networked Territories», «Innovative Territories» and «Low-Density Territories». Such territorial configurations reveal the intertwined interpenetration between social inequalities and the main dynamics that today run through contemporary societies. European Union would benefit by considering these territorial inequalities for achieving social cohesion.

Keywords: territorial inequalities; local system of indicators; place-based approach; European statistics; inequality paradigm.

Title of abstract: Quality assessment of the Gender Equality Index

Davide Barbieri, Anna Fiore, Irene Rioboo – European Institute for Gender Equality (EIGE), Vilnius, Lithuania

The Gender Equality Index is a composite indicator built by the European Institute for Gender Equality to measure the progress of gender equality in the EU Member States over time. It is formed by 31 indicators divided into six core domains (work, money, knowledge, time, power, health). Additionally, it pays attention to intersecting inequalities and violence against women. Launched in 2013 for the first time, the Gender Equality Index has reached the 6th release in 2021.

One of the main criticisms, addressed to composite indicators is the high degree of subjectivity frequently present in their computation process. The Gender Equality Index overcomes this limitation by applying a rigorous quality assessment in each of its calculation steps.

Starting from the raw sources used, the selection of variables is done from the main official statistical sources at European level, applying both conceptual and statistical criteria that guarantee the data quality in terms of relevance, accuracy, comparability and geographical coverage. In a following step, the use of multivariate analysis warrants the internal coherence of the data, keeping only relevant variables that group together based on their internal relationship and creating the conceptual structure of domains and subdomains. Then, a robustness analysis is carried out to select the best formula to aggregate the variables and to compute the Gender Equality Index, meaning the most stable formula in terms of scores and ranks of the Index. Once the Index is computed, an additional check of its internal coherence is run, analysing the contribution of each variable to the domains, subdomains and global Index. Finally, an uncertainty and sensitivity analysis are done to both quantify the uncertainty in the ranks of the Index as a result of the different decisions taken along its computation, and quantify the uncertainty caused by each individual assumption.

This solid methodology applied in the development of the Gender Equality Index is based on sound statistical principles internationally agreed, assuring its quality and integrity.

Keywords: Composite indicator; robustness; uncertainty; sensitivity.

Title of abstract: Preparing the European Statistical System for the Use of Privately Held Data

Peter Struijs, Albrecht Wirthmann, Fabio Ricciato, Konstantinos Giannakouris – Eurostat

In respect of data sources, official statistics rely mainly on surveys and administrative records. The environment of official statistics, however, is changing rapidly, leading to opportunities related to the potential use of new data sources. These opportunities need to be grasped, if only because solely relying on the more traditional data sources is becoming increasingly unsustainable. Examples of highly relevant data from private data sources are financial transactions data, mobile networks data, sensor data, data from internet platforms, and data related to the Internet of Things. Using such data would substantially increase the quality of statistics in all its dimensions, in particular relevance (e.g. statistics on new phenomena such as the collaborative economy), timeliness (near real-time statistics may become feasible if surveys are augmented with data from new sources) and coverage (new data sources allow for high levels of detail). Moreover, it creates opportunities for reducing costs and response burden.

Therefore, the ESS has been trying, in various ways, to prepare itself for the integration of new data sources in the production of official statistics. This had led to some significant results, such as the use of scanner data for price statistics, but there is still a long way to go. The paper provides an overview of what has been achieved so far, what is currently being undertaken, and what lies ahead. Specifically, the following subjects will be covered by the paper: the work of relevant Commission expert groups (in particular the one concerned with business-to-government data sharing for official statistics), legislative initiatives (in particular the forthcoming Data Act and the envisaged revision of the European Statistical Law), the internal preparation by the ESS (in particular the position paper of the ESS Committee on the Data Act, the work of various ESS bodies and groups, and the attempts to build an adequate infrastructure for the future), and the way stakeholders of all these efforts are engaged.

The current developments will determine the future shape of the ESS. The authors would like to share their ideas with the participants of the Q2022 Conference and give them the opportunity to influence these developments by giving feedback.

Keywords: New Data Sources, Data Access, Privately Held Data, Statistical Legislation

Title of abstract: Official statistics – Q(uo) vadis?

Hans Viggo Sæbø, Marit Hoel – Statistics Norway, Oslo, Norway

The data revolution has resulted in discussions in the statistical community on the future of official statistics. Will official statistics survive as a brand, or will such statistics drown in the flow of data and statistics from new sources and actors, including misused statistics and fake news? Will statistics survive as a concept in a world driven by data? The Covid-19 pandemic has been an additional driver for discussion. There is a need to maintain the quality of official statistics and highlight the value of such statistics for the users as a basis for – and supplement to – other statistics and information. It is at the same time important to implement new developments to improve and keep up the relevance of official statistics.

Key pillars today are statistical legislation, quality frameworks and core values defining requirements to official statistics. Possibilities are also linked to new statistics, use of new data sources and possible extended roles of the statistical institutes within coordination, collaboration and data stewardship. The paper addresses these issues in the light of trends in official statistics since the UN Fundamental Principles of Official Statistics were formulated about 30 years ago. Quality challenges for statistics and dilemmas in defining the roles of statistical institutes are considered.

The paper considers the potential for official statistics to develop its role and position, for maintaining trust and making facts matter. The paper includes examples from Statistics Norway that are believed to be relevant for several statistical institutes.

Keywords: official statistics; data revolution; coordination; cooperation; data stewardship.

Title of abstract: Principles for the Use of Privately Held Data for Official Statistics

Peter Struijs, Albrecht Wirthmann – Eurostat

In view of the rapidly changing environment of official statistics, solely relying on data from surveys and administrative sources is not sustainable anymore. Traditional data sources should be augmented by new types of highly relevant data, such as mobile networks data, sensor data, data from internet platforms, or data related to the Internet of Things. Such data is needed for describing new phenomena like the collaborative economy or new forms of labour, for increasing timeliness and the level of detail, and last but not least for increasing the quality of statistics, reducing costs and response burden. The Bucharest Memorandum on Official Statistics in a Datafied Society coined the term Trusted Smart Statistics for this concept.

To a large extent the new types of data are exclusively held by private organizations. So-called Business to Government (B2G) data sharing mechanisms need to be arranged to tap these data. A B2G expert group, formed by the European Commission, has provided recommendations for B2G data sharing for public interest purposes in general. Another Commission expert group is currently specifically looking at B2G for official statistics. This is the so-called B2G4S expert group, which is expected to produce their final report in early 2022. Moreover, Eurostat has initiated a Group on Using Privately Held Data for Official Statistics, which is also looking at the needs and conditions for Official Statistics. The Eurostat group is trying to link as much as possible to the findings and recommendations of the expert groups at European Commission level.

One of the main products of the Eurostat group so far has been a list of ten main principles for the use of privately held data for official statistics. These principles, which have been endorsed by the ESS Committee, range from the application of professional standards to adequately respecting businesses interests, from not asking for more data than minimally needed to creating a level playing field for the data providers, from choosing the right access modalities to applying a well-considered approach to the issue of costs. Similar principles aiming at data sharing on a voluntary basis have been formulated by other groups and committees before. In contrast to this, the principles formulated by the Eurostat group aim at providing a basis for the ongoing revision of the legislative framework, aimed at facilitating the (re)use of privately held data for official statistics.

Apart from presenting the proposed principles, the authors want to give an opportunity for feedback from the participants of the Q2022 Conference, as gaining access to and using privately held data will be crucial to the Trusted Smart Statistics concept of the future.

Keywords: Big Data, Data Access, Privately Held Data, Statistical Legislation.

Title of abstract: Data stewardship: To be, or not to be, that is the question

Ana Carmen Saura, Yolanda Gómez, Ana Cánovas – INE, Madrid, Spain

In our process to adapt official statistics to the new data ecosystem the debate on the new possible roles of the national statistical offices burst onto the scene and, as a consequence, the idea of "data stewardship" is on everyone's lips at present. Some National Statistical Institutes (NSIs) have already become national data stewards, other have gained competences on data management and design, others are key advisers inside their governments...but, in general, many questions are still on the air, what, when, how and, in particular, how far.

In this article we will analyse the different approaches towards data steward from the perspective of a National Statistical Institute, detailing the advantages and disadvantages of each of them. The different models of data governance will also be considered, with especial focus on the role of NSIs, the competences that could be assumed according to the Data Strategy and how we could manage to do it.

In addition, a common solution inside the ESS to move forward data stewardship will be presented, including new provisions in Regulation 223/2009 on European Statistics to somehow harmonize possible new roles.

Keywords: Data stewardship; Data strategy; Data governance.

Title of abstract: The role of the National Statistical Institutes in a new European Data Framework

Marius Andersen – EFTA Statistical Office, Luxembourg

The European Data Strategy aims to create a single market for data by facilitating easy share and use of data across sectors, all within a new data governance framework. The European data strategy will therefore define how Europe will work with and use data in the future and define the actors by creating a new legal framework for data.

The legal framework created by the European data strategy will define data sharing mechanisms, include several types of data and define data sharing between several types of actors. Business to government data sharing is perhaps of the most obvious interest to the National Statistical Institutes. Access to privately held data has been a long-standing strategic objective for National Statistical Institutes with its expected potential to improve access to and use for official statistics.

In parallel with developing a new data framework, discussions on data stewardship have already been ongoing for years in the global statistical community. Cleary, this discussion together with the initiatives concerning European data raises a series of opportunities and challenges for the National Statistical Institutes and partners in the European Statistical System.

Based on the experiences made from the conference of the EFTA Statistical Office on data stewardship held in December 2021, this paper considers the different roles and functions a National Statistical Institute can take as a data steward. Furthermore, the paper discusses possible interlinkages between the framework created by the European data strategy, data steward functions, and considers the opportunities for national statistical institutes in this framework. While the data steward role might not be entirely new to the National Statistical Institutes by now, a new data framework in Europe is likely to enable the National Statistical Institutes to look further into their role as data stewards.

Keywords: data stewardship, European data strategy.

Session 38
NEW CAPABILITIES AND PRACTICES FOR OPEN AND BIG DATA TO PRODUCE STATISTICS

Title of abstract: Quality improvement in mobile phone position data: A collaboration model between SCB and an operator

Pieter Vlag, Jens Malmros, Ulf Durnell – SCB, Sweden

Anonymised and aggregated mobile phone position data from mobile network operators can be a very suitable data-source for producing official statistics about dynamic populations, commuting, tourism and long distance travelling. Having this in mind, Statistics Sweden (SCB) and one of the biggest telecom operators in the Nordics (Telia) started a collaboration in 2020, one month before the outbreak of COVID-19. The aim of the project was to analyse the quality of aggregated mobile phone data.

Because the data clearly showed changes in population movements in relation to the pandemic, public interest and interest from authorities grew considerably after the outbreak of the pandemic. At the same time, it appeared that creating stable time-series out of mobile phone position data is not straightforward in periods of sudden change, hampering an exact quantification of differences in population movements before and after the pandemic.

Transparent and well-described processes were needed to tackle this issue. In this context, the collaboration gradually changed into a partnership in which the operator delivered 1) process descriptions and 2) aggregates estimated with different parametrisations and SCB developed 1) quality reports and 2) benchmark tables. Benchmark tables compare differences between dynamic population estimates based on mobile phone position data with register-based static populations from SCB and related register-based statistical data from SCB.

The results of this partnership will be discussed in the presentation. The results of the next and still on-going phase of the partnership will also be presented: development of statistics out of the partnership. The involvement of other mobile network operators in this process will also be discussed.

Keywords: mobile phone position data, time-series, benchmarking, quality reporting, partnerships.

Title of abstract: Using webdata to derive the economic activity of businesses

Johannes Gussenbauer, Alexander Kowarik – Statistik Austria, Vienna, Austria
Eline Toledo – former Master student, univ. Leiden
Arnout van Delden, Dick Windmeijer – Statistics Netherlands, the Hague

The economic activity of businesses is often a key characteristic for the production of most business statistics. This characteristic is determined for all units in the population and stored in the statistical business register. Determining and maintaining the main economic activity of a business is a difficult classification task which often needs manual editing. This is a very time-consuming task, especially for national statistical institutes with a large businesses' population size. Misclassifications in the economic activity can lead to seriously biased business statistics estimates.

Due to an increasing web presence of businesses, the use of data on the web is starting to become a viable data source to help to classify the economic activity of businesses. One project within the ESSNet Web Intelligence Network, that started in April 2021, aims to develop automatic procedures to support manual editing of business registers NACE codes. That is done by predicting the economic activity using business website texts (and possibly other information) which can subsequently be used to find potential errors in the currently registered codes. In this paper we present some first results in the form of two case studies.

The first case study, from Statistics Austria, presents the current state of an automated procedure to predict economic activity on NACE-2 level. The procedure starts with identifying webpages of businesses in a statistical business register and scraping relevant text from these webpages. Afterwards the text is processed using natural language processing techniques and finally used in a neural network to predict the most likely NACE codes.

The second case study, from Statistics Netherlands, explores the use of knowledge-based features on predicting economic activities in section R (Arts, entertainment and recreation) and in some related NACE codes. Different forms of using knowledge-based features are compared with a classical bag-of-words approach, for a naïve bayes classifier, a support vector machine and a logistic model. In the near future, within the ESSnet project, we hope to compare the model performance of different text mining approaches, like the effect of different types of features, different machine learning algorithms and different types of data sources.

Keywords: Machine Learning; Natural Language Processing; New Data Sources; Business Statistics.

Title of abstract: Development of methodology for automated crop mapping in Greece using Neural Networks and Sentinel-2 satellite imagery

Eleni Papadopoulou – Department of Mathematics of the Aristotle University of Thessaloniki, Thessaloniki, Greece

Reliable and accurate cropland mapping at national, regional and global level is essential within the framework of Sustainable Development, food security and sustainable environmental management. Earth Observation data, collected by satellite sensors of European Union's Copernicus Earth observation programme, provide the opportunity for the development of automated high accurate methods for information extraction at regular temporal intervals.

Within the framework of the current study, computational approaches for cropland mapping in Greece using deep learning methods and high spatial resolution Sentinel-2 satellite images of the Copernicus programme were developed and evaluated. The study area was selected within the Regional Unit of Serres where reliable reference data from OPEKEPE (Greek Payment Authority of Common Agricultural Policy (C.A.P.) Aid Schemes) were collected for a complex classification scheme consisting of 20 classes. The experimental design included five supervised machine learning classification algorithms, including a Temporal Convolutional Neural Network, a bidirectional Gated Recurrent Unit Neural Network, a bidirectional Long Short-Term Memory Neural Network, a combination of a Recurrent and a 2-D Convolutional Neural Network and finally a Random Forest. Furthermore, the effect of different methods of training sample extraction was examined and evaluated, not only in terms of classification accuracy metrics, but also considering classification uncertainty.

The results of this study, as indicated by the high classification accuracy (more than 90%), support the transfer of the approach nationwide, for the operational inventory of spatial explicit information over agricultural areas.

– This study was carried out in the frame of MSc thesis in Mathematics with specialization in "Statistics and Modelling" which is directly associated with EMOS program.

Keywords: remote sensing; land cover mapping; crop classification; uncertainty.

Title of abstract: Data science skills for the next generation of statisticians

Dario Buono – Eurostat, Luxembourg, Luxembourg

The capability to incorporate new data sources and to benefit from disruptive technologies will be at the core of the modernization of official statistics. In short, new data sources are a way to meet users' needs better. One key factor in meeting these challenges is the development and building of the necessary skills and competences. In a nutshell, what are the competences, skills and features we are looking for the next generation of statisticians, and 2) how can these competences and skills be acquired? Statistical organisations will have to create favourable conditions for new production methods and using data science skills with success. This means, for example, 1) establishing an innovative culture, where experimental activities are commonplace, 2) building and maintaining collaborative and multidisciplinary data science teams, 3) recruiting individual data scientists and using long-term personnel planning, 4) training and supporting personnel in identifying themselves with new competence requirements, and 5) aspiring to management and leadership practices, which make these changes possible.

Keywords: Data Science, New data sources, Innovative Methods.

This work is a revised version of a paper presented at CES 2017.

Speed talk session 1
PROCESSES AND STANDARDS

Title of abstract: Cost Per Process Model

Nuria Torroja Mateu – Instituto Nacional de Estadística y Geografía (INEGI), Ciudad de México, México

Identify direct and indirect production costs, as well as indirect institutional costs, for each phase of the statistical and geographic information production processes. The Model used to identify costs per process is based on the Quality Indicators for the Generic Statistical Business Process Model (GSBPM). The expected results in this measurement consist of having a better budget planning in the allocation of financial resources to statistical or geographic processes, as well as to cross-cutting activities.

Keywords: Model, Costs, Process

Title of abstract: Standardization process of budget administrative data, having statistic characteristics and usefulness for public finance control

Daniela Collesi – Italian Ministry of economy and finance, Rome, Italy

The accounting reform introduced in Italy with Law 196 of 2009 has, among its various objectives, the harmonization of accounting. It defines a series of rules that make possible to produce data ready to be used both for the administrative purposes of government and for production of public finance statistics. To achieve this goal, it was decided to adopt, in the administrative area of public budgets – for all phases of preparation, management and reporting – the classifications of official statistics, according to the rules of the national accounts.

The introduction of new classifications concerning both budget schemes and the operations of the management of public budgets during the year, is significant because it determines a real earthquake for the administrative structures of the budget, which very often do not have a suitable professional training oriented to statistics and classifications; all this is, of course, apart from the fundamental objective of having quality data useful for a correct evaluation of public policies, making the data homogeneous and therefore comparable between different public bodies and consolidable between them.

To respond to the need to support administrations in their accounting revolution, a four-year experimental process was launched for local authorities – regions, provinces, municipalities and their subsidiaries bodies – which make up the majority of Italian public bodies. This process has enabled the tools available to them to be refined, including through a subsequent permanent commission that analyzes the main critical issues, proposing common shared solutions aimed at ensuring uniformity of the rules for all the bodies falling within the scope considered. For expenditure, which provides for identification of the purpose through the definition of Missions and programs, associated with the international COFOG classification, a glossary was prepared that defines in detail the content of each program and the corresponding COFOG.

The acquisition of budget data and their analysis for the five-year period 2016–2020 shows interesting results that confirm both the importance of the direct adoption in the budget of the standard classifications of the national accounts and the need for support for the entities and their ability to improve over time, as evidenced by the increasing quality of the data in the last two years, compared to the 2016 data, the first year in which the accounting reform was extended to all entities in addition to those that had experimented. The next step will be to make the overall result of the analysis available to the entities and their associations so that they can appreciate the characteristics of increasing homogeneity over time and the possibility of comparison between them, both in time and space.

Keywords: first keyword budget data; cross-institutional comparability; second keyword: new budget rules; third keyword: adoption of statistical classifications.

Title of abstract: SDMX XML files production and dissemination: Burden or Opportunity?

Sónia Quaresma, Pedro Cunha – INE, Lisbon, Portugal

Data is exchanged between the institutions, including NSIs and Eurostat in pre-agreed formats. Nowadays this usually means SDMX ML files. Despite all the advantages that this format conveys it is not so simple to produce as an excel or csv file. In fact, in many cases, like the National Accounts domain that will be presented in the current paper, an intermediary format is used – excel template files. These templates are easier to fill by domain experts and don't require an IT background. However, these templates may change or even be phased out as is the case in NAMAIN (National Accounts Main Tables). In this case to be able to continue producing the SDMX files the domain experts have to resort to the IT department help but they lose some independence in the process.

Something as simple as producing files in the correct format becomes a burden. When the complexity should be in producing the content, it shifts to shaping the output but this is not the only way. Through modernization of the whole statistical production process, it is possible to guarantee all the necessary products, enhance the quality of the products by encompassing more validation in the processes and promote the domain expert's independency for production and analysis.

The current paper explains how this was implemented for NAMAIN domain at INE Portugal, with an interdisciplinary team that streamlined the process from beginning to end. The approach followed includes some changes to the usual processes:

– The statistical data warehouse (SDW) was designed together with the domain experts to ensure that they would be able to load the data themselves, instead of the IT as is generally the case.

– Additional parameterization and redundancy were introduced so that the domain expert users could recover from problems that happened during the data ingestion, and not be dependent on the IT intervention.

–   The step of data loading was enlarged to constitute a de facto data archiving, meaning that the data is simultaneously made available for the current use, like exploration, validation or delivery and archived in a versioning system.

–   The validation procedure was built using the validation template for the main types of rules which is able to execute Eurostat's validation rules in SQL (database querying language) in the data warehouse and present the results in tabular form (Excel or Power BI).

–   SDMX RI was coupled with the SDW and the required outputs were mapped to its objects so that the files can be generated by the domain expert users through a graphical interface at the press of a button.

These 5 steps are described in the next sections of the paper.

Keywords: SDMX XML files, Statistical data warehouse; Validation; Modernization of Statistical Production Processes, NAMAIN

Title of abstract: Reproducible analytical pipelines: software engineering in analysis

Alexander Newton – Office for National Statistics, London, United Kingdom

Reproducible Analytical Pipelines (RAP) are the application of good practices from software engineering in analysis and research. They make analysis and research more reproducible, better quality, more trustworthy, and efficient. The recent Office for Stats Regulation review of RAP says "there is a huge potential to apply RAP principles both in and beyond official statistics".

RAP is the linchpin of the digital transformation of government analysis. Through the high quality automation it brings, more effort can be directed towards building the research products that our users need. Recent successes like the UK Coronavirus Dashboard show how powerful government data can be in answering the needs of the public when sound RAP principles are applied.

However, learning the practices in RAP represent a substantial challenge for many government analysts – especially in a context where there is little prioritisation to build this capability and where the tools needed can be difficult to acquire. Our Coding in Analysis and Research survey shows that although most analysts feel it's important to implement the RAP principles in their work, many fewer are able to do so.

Our approach to RAP is multi-pronged to tackle the cultural, capability, and technical barriers that exist.

We operate a Champions network that engages across government and delivers mentoring, training, and advocacy within their respective departments,

We deliver consultancy and mentorship for organisations beginning their journey,

We advocate for the digital transformation of analysis with RAP at senior levels in government,

We monitor progress towards RAP and evaluate interventions,

We role model good practice in our own analytical work.

Through our research of coding in analysis across government and the research of colleagues we identify next steps for public sector organisations and advocate for the inclusion of RAP principles at the heart of government analytical strategy.

Keywords: reproducibility, unit-testing, data-engineering, software engineering, culture change

Title of abstract: Efficient quality management tools, Self-assessment

Vugar Mammadalizade – The State Statistical Committee, Baku, Azerbaijan

In the modern era, where the importance of official statistics is growing, the application of standards on quality management and self-assessment tools adopted in the international arena is of a great importance in terms of organization of effective management in providing users with quality statistical.

In official statistics, there are many standards adopted in the international arena and implemented in different countries by various institutions and organizations as ISO 9001 Quality Management System, a model of improvement of the European Foundation for Quality Management (EFQM), Common Assessment Framework (CAF) which includes the requirements of the International Standardization Organization as efficient tools of the quality management system. Since statistical data accuracy, quality and compliance with social-economic events taking place in the society is of a great importance, the quality assurance in official statistical system of Azerbaijan as well as in the developed countries of the world for quality assurance of administrative and statistical processes organized and statistical products developed in the state statistical bodies has become one of the relevant issues of the recent period.

Self-assessment, which is one of the main principles of the quality assurance, is one of the most important and efficient tools applied by statistical agencies for improvement of the statistical system in recent years. Identifying potential discrepancies and taking timely preventive measures are based on quality assurance. Many leading countries of the world have achieved the improvement of the statistical system by implementing self-assessment in their systems and, as a result, increasing the confidence of users in statistics, as well as the image of the statistical agency among users. Self-assessment in the statistical system is carried out through the national basic principles of quality assurance, the European Statistics Code of Practice, Self-Assessment Questionnaire for statisticians – DESAP self-assessment and internal audits of the quality management system.

The State Statistical Committee of the Republic of Azerbaijan is taking appropriate steps towards continuous improvement of its activity. Internal audits of the quality management system enable efficient assessment of each statistical business process by performing the function of monitoring the work done in the field of quality assurance.

Keywords: ISO 9001; quality management; DESAP; self-assessment; internal audit.

Title of abstract: The commitment to quality in regional statistical production: some evidence from Catalonia's Statistical System

*Xavier Cuadras-Morató, Cristina Rovira, Carme Saborit, Enric Ripoll, Jordi Galter –
Statistical Institute of Catalonia, Barcelona, Spain*

The Statistical System of Catalonia provides information for its territory, which is entrusted by the respective legal standards (autonomy statute, statistical laws, statistics plans and programs). The elaboration of the next Statistical Plan of Catalonia 2023–2028 (SPC) aims to include, among other objectives, the latest changes and innovations in official statistics such as the new sources of emerging data, the use of new technologies and to update the legal framework to reinforce the quality of the System in compliance with the European Statistics Code of Practice.

As part of the preparatory work for the SPC, the Statistical Institute of Catalonia set up, among other tasks, the analysis from a regional point of view of the degree of alignment of the Catalan official statistics with the EU statistical regulatory framework, and a strategic review of the official statistical production to respond to future challenges that included a structured consultation addressed to users from both inside and outside the Catalan Public Administration.

The goal of this research was to evaluate the quality of the current statistics production and make proposals that define the vision for the direction of statistics production for the next years in a context where the economy and society are changing more and more rapidly. Additional challenges included the detection of new emerging statistics, the enhancement of conceptual and territorial detail, the promotion of experimental statistics or the use of new data sources.

To carry out the first of the two tasks, Eurostat's Statistical Requirements Compendium, a summary of European statistical production, was considered. The official regional statistics that were not aligned with the EU regulations were identified as well as the correspondent relevant issues and variables in the EU regulations that are not covered (or partially covered).

Assessing the quality of official statistics, a public good, requires measuring its multiple dimensions. Therefore, it is essential to rely on the assessment of its users. To obtain this valuable information, that provided a broad external assessment of the content of Catalonia's Statistical System, a structured consultation was carried out with the participation of more than 130 different types of users who had to evaluate each statistic and provide advice regarding the identification and prioritization of possible new contents to be included in the next Statistical Plan.

The paper describes the work carried out, the main conclusions reached and the specific proposals that, based on these two main tasks, have been included in the draft of the new SPC which is expected to take effect in 2023.

Title of abstract: Data web apps as a tool for promoting communication and dissemination of official statistics

Stratos Moschidis, Christina Karamichalakou, Athanasios C. Thanopoulos – Hellenic Statistical Authority, Piraeus, Greece

Official statistics place particular emphasis on communication and dissemination of surveys results to citizens and stakeholders. This is typically done through the publication of press releases and presentation of aggregated data of statistical surveys. The use of web services and software tools that allow users to interact with the results of official statistics comes to further enhance their dissemination. This paper explains how this functionality can help to broaden the communication and understanding of official statistics while presenting the methodology for implementing such an online application without the specialized knowledge of frontend backend programming techniques. For demonstration purposes, an experimental version of such an application was constructed that presents in an interactive way the quarterly results of the new statistical product of the Hellenic Statistical Authority (ELSTAT) on Greek business demography.

Title of abstract: Quality Assessment of Official Statistics in TurkStat

Fatma Deniz Özkan, Bilge Çorumlu, Serdar Cihat Gören – Turkish Statistical Institute, Ankara, Turkey

National Statistical System in Turkey is a multi-actor system. Turkish Statistical Institute (TurkStat) and Other National Authorities (ONAs) take place within this system. TurkStat consists of a central office and 26 provincial offices. It was structured as the Presidency of Turkish Statistical Institute and The Statistics Council, based on the Statistics Law of Turkey No 5429. The Statistics Council is established to advice on the development and implementation of the Official Statistics Programme (OSP).

Official Statistics Programme (OSP) ensures the effective functioning of Turkish Statistical System. As the coordinator of Turkish Statistical System, TurkStat has a mission to undertake quality reviews of official statistics which are produced within the framework of the Official Statistical Programme.

TurkStat has been doing quality review, also named as "Quality Logo", of official statistics produced by ONAs since 2015. Quality logo is a system in which official statistics are assessed against national quality principles based on European Statistics Code of Practice (ES-CoP). As a result of these assessments, improvement actions are defined and regularly monitored. These reviews not only improve the quality of official statistics but also it strengthens the cooperation and coordination role of TurkStat.

Output quality of statistics produced by TurkStat is monitored through producer and user-oriented quality reports, and other quality assurance methods applied during the production process. In order to assure process and output quality and ensure compliance with ES-CoP principles TurkStat developed a new regular quality review system for official statistics produced by TurkStat. For this purpose, a tool called "Quality Monitoring and Assessment Tool" ("QMAT") has been developed in TurkStat in order to apply the quality assurance, to measure the potential sources of error at each stage of the statistical production process. QMAT consists of a self assessment questionnaire, list of required documentation and a plan to review the quality of statistics produced by TurkStat. A committee has been formed to review the questionnaire and documentation, to conduct peer review like meetings with the subject matter departments and to write a report summarizing the findings of the quality review and proposed improvement actions.

This paper summarizes both theoretical and methodological information about the quality review system of TurkStat (quality logo and QMAT), some key takeaways from the quality review process.

Keywords: Turkish Statistical System, official statistics, quality review, quality assessment, quality logo.

Title of abstract: Quality of Administrative Data for Statistics: Developing a User-Friendly Quality Framework

Kathryn Reynolds, Ella Williams Davies – Office for National Statistics, Newport, United Kingdom

Administrative data is data that has been collected during the operation of an organisation, e.g. hospital records, benefits claims, financial transactions, etc. The UK Government produces an extremely large amount of administrative data that is potentially very useful for another purpose: producing statistics. However, as this purpose differs from that which the data were originally collected for, it gives rise to some unique and often complex challenges.

In order to effectively and responsibly use administrative data in this way, a solid and informed understanding of its quality and the resulting impact on the quality of outputs produced is essential. A number of frameworks have already been developed to assist with this, including Statistics New Zealand's (2016) "Guide to Reporting on Administrative Data Quality", and Statistics Netherlands's (2009) "Checklist for the

Quality Evaluation of Administrative Data Sources". However, due to the complex and varied needs of the users of statistics, and the complex and varied scenarios surrounding statistical production, a 'one size fits all' framework is difficult to produce. For example, the quality considerations involved when using a single dataset to produce estimates for England for one year differ considerably from those involved when combining multiple datasets to produce estimates over time at local authority level.

As part of this line of work, the Methodology and Quality Directorate at the Office for National Statistics have developed a quality framework. This will help those across the UK's Government Statistical Service working with these promising sources of data to make informed decisions about their quality and resulting use. Based around user needs and the idea of a proportionate assessment of quality (defined as 'fitness for purpose'), the framework has been shaped by user feedback at all stages of development. It aims to draw together much of the existing and valuable guidance in this area, collating useful tools and recommendations into one, user friendly resource.

This presentation will cover:

a) The user-focussed approach taken to develop the framework.

b) The challenges of framework development in such an evolving and complex area.

c) A brief walk-through of the framework itself.

d) Potential future plans.

Keywords: quality; framework; administrative data; statistics.

References

Statistics New Zealand (2016). Guide to reporting on administrative data quality. Retrieved from www.stats.govt.nz.

Title of abstract: Special Outlier Treatment Technique in Seasonal Adjustment for Handling the Effect of COVID-19

Gábor Lovics, Beáta Horváth, Mária Pécs – HCSO, Budapest, Hungary

In official statistics seasonal adjustment is often used to ensure the comparability of inter-annual data over time and space. During this process the effects of the outliers need to be managed. Outliers are real data – in the background there is an economic, social or environmental event – outside the expected range of values based on the general trend.

The type of outlier depends on time, extent and length of the effect. Estimating these properties can be a difficult task especially at the end of a time series. The effect of the COVID-19 challenges the experts of seasonal adjustment. On the one hand, the pandemic and the lockdown cause atypical time periods, when observations differ from the patterns of the other data of the time series. Therefore, outlier treatment could be a good tool to handle these effects. On the other hand, even now it is hard to decide for how long the COVID-19 has an influence on our time series.

The NSIs all over the world use different techniques to handle this situation. For example, some of them use outliers which connect to a real phenomenon that affects only one observation (additive outlier) in every time periods affected by COVID-19. Others use outliers which are linear or quadratic increase or decrease in the level of the series over a specified time interval (ramp or quadratic ramp outlier).

At HCSO outlier that begins with a great difference then it gradually decreases and at the end the difference disappearing (transitory change – TC) was often the chosen type. This outlier has a parameter (TC rate) which theoretically can take any value between 0 and 1. The parameter defines the rate of the gradually decreases. The smaller the parameter, the faster the time series returns to the initial level. In practice, this parameter is fixed constant 0.7.

During our analysis it was found that the effect of COVID-19 is often similar to what can be treated by TC, but the rate of the gradually decreases did not fit in every case. Therefore, an adaptive TC was used, where the parameter can be flexibly chosen in alignment with the time series. It was found that this small change makes it often more practical to handle the effect of COVID-19 in Hungarian time series.

Keywords: seasonal adjustment; time series; COVID-19; outliers; transitory change.

Title of abstract: Alternative methods to produce official statistics

Elsa Dhuli, Olsida Spiro – Institute of Statistics, Tirana, Albania

Covid-19 has been posing new challenges and use opportunities in the production chain of official statistics. In order to not interrupt the statistical production and meet user needs, INSTAT had to immediately react and adapt to continue the work considering the restrictions in the data collection process. In this regards, during the second quarter of 2020, different data collection techniques were adopted such as a combination of different data collection methods and new data sources were used. The quality principles in production of official statistics even in pandemic time were in focus of production, process and communication of statistical products.

This paper will present the actions undertaken and the changes made on the production of official statistics taking into consideration the limitations and increasing demands to produce qualitative and timely statistics. In order to evaluate the impact of these changes in the quality of the products, ESS quality and performance indicators will be used.

Although different data collection methods or new data sources were used, the quality dimensions for all statistical products have not been impacted.

Keywords: Covid-19, data collection, data sources, quality.

Title of abstract: Increasing challenge of non-response during the pandemic crises – how non-respondents influence the results of nation-wide survey

Maruša Rehberger, Darja Lavtar, Metka Zaletel, Nejc Berzelak – National institute of Public Health, Ljubljana, Slovenia

High response rate is besides other components of survey quality one of the crucial points of survey management, especially during the pandemic crisis when we want to reach vulnerable subpopulations. Therefore, we examined the differences between respondents and non-respondents in the nation-wide survey according to different sociodemographic groups.

In 2021, the Slovenian National Institute of Public Health (NIJZ), having a role of Other National Authority for health and health care statistics in the Slovenian statistical system, conducted cross-sectional Behavioural Insight Survey on Covid-19; National Survey on the Impact of the Pandemic of Life (SI-PANDA 2020/2021) in two waves. The target population were Slovenian residents aged 18 and older, among whom the probability sample of 16,000 persons has had been selected. Initially, the total sample was divided into two subsamples conducted in two consecutive data collections. In both waves, mixed mode design was employed with combination of computer-assisted web (CAWI) and paper and pen (PAPI) postal survey. The first wave has had been implemented from January 2021 until end of March 2021 and second wave from May 2021 until mid-July 2021.

The paper presents comparison of the response rates between different socio-demographic groups in both survey waves. Despite equal data collection methodology and solicitation procedures, response rate in the second data collection was 10 percentage points lower (38,5%), than in the first (48,9%), even though both waves were conducted in a short period of time. In the previous years, similar surveys achieved response rates higher than 50%, therefore current response rates need to be examined carefully to estimate the possible bias caused by nonresponse.

The aim of the paper is to investigate non-respondents according to their sociodemographic characteristics, like gender, age, marital status, completed level of education, statistical region and some other information available from the sampling frame. The second aim of the paper is to investigate respondents' characteristics according to their mode of participation (mail and web) and selected pandemic indicators. In-depth examination of differences in these characteristics between respondents and non-respondents allows identification of specific groups that are less likely to participate in the survey despite multiple contact attempts. Implications for non-response bias in overall estimates and segmentations are discussed. Paper therefore also investigates differences in characteristics of web and mail respondents to evaluate the effectiveness of using different modes in facilitating survey participation and reducing nonresponse bias.

Keywords: non-response; mixed-mode survey; survey quality; data collection.

Title of abstract: How to use a dashboard to assure quality in the data collection. A dashboard with data from our support channels

Herman Morken – Statistics Norway, Kongsvinger, Norway

This presentation focuses on how a dashboard can be a good tool to improve the knowledge and quality of the data collection. We made a dashboard with data from our support channels used by our respondents. The dashboard systematizes the contact our support channels have with survey respondents. It includes data on why the respondents contact our support services. Information from the dashboard gives important insight to everyone in the organization that find this information relevant for their work. I.e. those that either work with the support channels, specific surveys or assess the data quality. It is vital to have good information about the problems and feedback the survey respondents have on our surveys, that can be used to improve the quality on different aspects of the data collection. The dashboard gives easy access to systematised data for all the surveys for the last couple of years. The presentation will show the dashboard, how it was made, how it is used today and plans for further use.

Keywords: dashboard; data collection.

Speed talk session 4
DATA AND DATA QUALITY

Title of abstract: Quality Considerations for EIGE's Gender Statistics Database

Ligia Nobrega, Irene Rioboo – European Institute for Gender Equality, Vilnius, Lithuania

While much progress has been made to reach equality between women and men in the European Union (EU), remaining gender gaps still need to be addressed. Moreover, and owing to international conventions, gender sensitive policies are of prominent importance.

The adoption of a gender mainstreaming approach is therefore crucial to reach gender equality. It is a key action and thus should be effective in the tools used for the design, the implementation, the monitoring and the evaluation of policies. The follow-up of these goals is impossible without the evidences provided by statistics. This leads to an increase demand for gender statistics and need for additional guidance to improve the availability of data ensuring a proper mainstreaming of gender issues in all areas of statistical production.

Although many data are now available with a disaggregation by sex, the gender mainstreaming approach has not been totally applied in the statistical operations throughout the EU.

Gender statistics measure to what extent equality between women and men is achieved across all areas of life. Such statistics give countries the means to gauge their success in achieving gender equality against the dual benchmarks of other countries' achievements and their own histories. For gender statistics to serve this purpose, they must measure an aspect of gender equality consistently, accurately and in a timely manner. It is also necessary to ensure that the statistics are presented clearly and with sufficient accompanying information (metadata) to allow users to understand what has been measured and how the measurement has been carried out. These properties constitute the basis for the multidimensional concept known as quality of statistics. The European Institute for Gender Equality (EIGE) published in 2016 a Gender Statistics Database that is a one-stop source to gender statistics data facilitating structured and user-friendly access to gender statistics. Together with its Database, the agency also developed quality considerations applied to gender statistics included in the tool. Those considerations are tailor made both for users of the Database and for data providers. For users, the document provides evidence that quality considerations are at the core of the Database and offers the tools needed to understand the applicability and the limitations of the data available. For data producers and providers, the document furnishes recommendations and guidelines on how to produce data in a gender-sensitive way, avoiding common pitfalls that lead to gender biases, and provides information on how to document the quality of data and statistics, particularly focusing on the gender-relevant aspects.

Keywords: Gender statistics; quality; gender mainstreaming.

Title of abstract: Gender based violence as a cost for society

Alessandra Capobianchi, Isabella Corazziari, M. Giuseppina Muratore, Claudia Villante – Istat, Rome, Italy

Gender-based violence (GBV) is a long-standing drama, characterised by violent behaviours committed often within homes, where a person should feel safe and happy, where partners that should be instead the most trusted persons act it.

GBV is too often perceived mainly as an individual problem. It is a severe violation of human rights and a huge obstacle in the path towards equal opportunities between women and men.

GBV has instead consequences, not only at an individual level. Often GBV consequences cause lack of working days besides that of time for personal activities, due not only to injuries but also to fear, depression, anxiety caused by the violence. Often children are involved, and shelters are required to safe the woman and children lives. GBV consequences are indeed a cost for the whole society, a cost that drains more economic resources than what is commonly perceived.

The present study proposes a methodology to evaluate GBV economic costs, based on the definition of some women's life dimensions involved by the violence consequences. The dimensions can be tangible and not tangible. The methodology combines macro and micro perspectives, according to the available sources of data, to evaluate the economic costs of the defined dimensions. The main model is an accounting one: the average cost estimate for each consequence's dimension will be applied to the corresponding victims' consequence estimate from the national Women's safety Survey. The possibility to link individual records in the survey to other individual's data about for example health and psychological therapy costs, working status and temporary disability, legal costs and other individual information will be discussed, proposing the main countable model based on aggregated costs as an alternative when such linkage is not allowed.

Keywords: gender-based violence; social costs; data integration; accountability model.

Title of abstract: Usage of non-probability sample and scraped data to estimate proportions

Andrius Čiginas – Statistics Lithuania, Vilnius University, Vilnius, Lithuania

Vilma Nekrašaitė-Liegė – Statistics Lithuania, Vilnius Gediminas Technical University, Vilnius, Lithuania

Tomas Rudys – Statistics Lithuania, Vilnius, Lithuania

The proportions of enterprises owning websites are estimated using a probability sample in Information and Communication Technology survey. The data collected by a private company and those obtained through the web scraping at Statistics Lithuania provide two alternative data sources, but each of them suffers at least from the self-selection bias. The number of methods to adjust the bias and integrate auxiliary data with the survey data are examined. The integration leads to more efficient estimators than the estimators based only on the survey data. The accuracy measures of the estimators considered are evaluated as well.

Keywords: Big data; coverage bias; post-stratification; calibration weighting; accuracy estimation.

Title of abstract: Digital cooperation on survey implantation in Tanzania and Mozambique aiming to observe data quality in all aspects

Mr. Dag Roll-Hansen, Mr. Kristian Lønø – Statistics Norway (SSB), Oslo, Norway

Mr. Titus Mwisomba, Mr. Samwel Kundael Kawa – National Bureau of Statistics (NBS), Dodoma, Tanzania

Ms. Sandra Maria António Caetano Dias Pinoca – Instituto Nacional de Estatística (INE), Maputo, Mozambique

We are facing a future where the quality and possibilities inherent in electronic communication enable us to cooperate across borders without crossing these borders physically. This is good new as resource constraints and the climate crisis forces us to rethink the necessity of travelling.

On this backdrop the National Bureau of Statistics (NBS) in Tanzania, Instituto Nacional de Estatística (INE) in Mozambique and Statistics Norway (SSB) embarked on large-scale survey project; a survey on access to sustainable energy and cooking stoves was to be conducted both in Tanzania and Mozambique. The survey is called Impact of Access to Sustainable Energy Survey (IASES) and aims to give quality data needed to address issues of planning, policy design and decision making at regional and country level as well as data to inform Sustainable Development Goal (SDG) number 7. The survey results will give guidance on how governments and partners can face challenges associated with expanding the electricity grid, understanding the impact of access to sustainable energy and achieving SDG 7. Extensive face-to-face-cooperation was planned when the pandemic put work on hold, but use of new technology helped us to accomplish our joint project.

The paper discusses how we overcame the challenges, creating new cost-efficient modalities of cooperation that reduces the need for travels, while applying international statistical standards to maintain the required data quality. This includes, but is not limited to:

– Use of e-learning, including a course on sustainable energy from the World Bank's Open Learning Campus and a YouTube channel with introductions and video lectures;
– Use of GitHub for update and exchange of code, including initial development, updates and debugging; and
– Use of Zoom and WhatsApp for regular and rapid communication.

The paper describes new applied approaches towards implementation of the joint survey on sustainable energy, aiming to observe data quality in all aspects. It also has a section on lessons learned and how digital cooperation on future surveys can be improved to enhance the quality of data from applied surveys.

Keywords: digital cooperation, survey implementation, e-Learning, sustainable energy statistics.

Title of abstract: The role of official statistics as a trusted source of information in a datafied society

Natasha Cenova – State Statistical Office, Skopje, Republic of North Macedonia

This paper analyses the role of official statistics and its relevance in the context of usage as a trusted source. One of the main goals of the official statistics is to serve to the public needs. Today in the digital society not every data is statistical data. Data on construction statistics in our SSO are very widely used in the public, private and government sector, since they provide important information on the economy and monetary policy as well as for future investments in the construction sector. The indicators for issued building permits and production in construction are part of the PEEI main economic indicators for the economic developments in the EU and are collected according to the Eurostat Methodology of short-term business statistics. In 2014, the Government of the Republic of North Macedonia implemented online data collection via an electronic system for issuing building permits, e-permits, prepared by the responsible authorities – the Association of the Units of Local Self-Government (ZELS) together with the Ministry of Transport and Communication. In cooperation with these institutions, the State Statistical Office incorporated in this process the statistical variables needed for the survey. In the Agency for Real Estate Cadastre (AREC) established and maintained is a Register for Prices and leases of the real estate. They publish a quarterly report where they implement and use our statistical data on issued building permits, finished and unfinished dwellings. This proves that they are trusted as a valuable source. The statistical data for this real estate market is very proactively used by local self-government; profit and non-profit organisations focussed on environmental protection. This proves that in this datafied society they are used as a trusted source among other data, and in the same time we shall focus and influence to other administrative sources that may affect statistical use.

Keywords: official statistics; building permits; PEEI; AREC, ZELS.

Title of abstract: Data ecosystem for the improvement of official statistics: example of a labour statistical information system

Carme Saborit, Núria Bové, Cristina Rovira – Statistical Institute of Catalonia, Barcelona, Spain

The Labour Statistics Information System (LSIS) of the Statistical Institute of Catalonia is designed as a data ecosystem whose main goal is the efficient and enriched production of statistics related to the labour market. It is based on the great potential of administrative registers to generate good quality information. It is also composed of other sample base sources and allows for the use of other internal and external sources. Also, new possible sources could be added in the future.

The core of the registry base is made up of the information from the Social Security affiliations and contribution accounts, as well as other records of administrative origin such as job seekers, contracts or those affiliated with other systems. It is interconnected with the population, entities and territory statistical registers, and their respective satellite registers (education, social benefits, etc.), which allow for the analysis of the information from different perspectives and in a common environment that provides greater coherence and quality to the data, with full respect for the protection of personal data.

The goals of the system are multiple. On the one hand, results from different point of views (population, business or territorial) can be obtained. They are also available in a transverse or longitudinal form. At the same time, it constitutes an efficient storage system allowing for its permanent update, which reduces production costs and shortens the time for preparing statistics, indicators, and related products.

In this paper, we review the main goals and the conceptual scheme of the system. We also provide a series of examples of the new statistical information that the system is currently producing and of the future statistical products waiting in the pipeline for the next future.

Keywords: quality; labour statistics; integrated system; conceptual scheme.

Title of abstract: Evolution of survey question wording: mixed mode cognitive testing for a mixed mode survey

Maruša Rehberger, Darja Lavtar, Vito Klopčič – National institute of Public Health, Ljubljana, Slovenia

Gaja Zager Kocjan – University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia

Providers of official statistics increasingly recognize a critical role that qualitative pre-field testing has in the processes of development of new questions and translation and national adaptation of the existing questions. Pre-field methods include cognitive interviews (CIs) that focus on investigating mental processes involved in answering survey questions.

Slovenian National Institute of Public Health, in the Slovenian statistical system holding a role of the Other National Authority (ONA) for health and health care statistics, conducted qualitative pre-testing of the Slovenian version of new questions that were to be included in the 3rd wave of the European Health Interview Survey (EHIS): oral health, dietary habits, physical activity, use of electronic cigarettes, etc. The original English version of the new EHIS 2019 questions already underwent cognitive testing, but the national cognitive testing was deemed necessary as well in order to find the most appropriate Slovenian translation of the questions, answer categories and explanatory texts.

The aim was to adjust and improve translated questions through means of cognitive testing for a mixed-mode data collection method (web survey and face-to-face interviews were used in EHIS 2019 data-collection). A two-step approach was employed: (1) first, face-to-face CIs with more general probes were conducted on an initial smaller sample of participants, following by (2) web probing with focused probes on a larger sample of participants. Issues that remained open after face-to-face CIs were investigated through web probing (e.g. examples of what should be included or excluded when answering the specific question). A web questionnaire was developed with targeted, close-ended probes that were easy to understand and answer. Some probes were focused specifically on the usability of the web environment of the OneClick survey (1ka.si) tool which was used for the web testing and for the EHIS 2019 survey itself. The benefits of applying a two-step mixed-mode cognitive testing include (1) collecting a larger amount of feedback from different individuals, (2) probes focused on greater details, and (3) saving time and money. The evolution of some survey questions from their original version to the final wording included in the EHIS 2019 will be presented.

This two-step approach can also be adjusted in pandemic times: face-to-face CIs can be conducted with video conference tools.

Keywords: qualitative pre-testing, cognitive interviews, web probing, question development, mixed-mode pre-testing.

Title of abstract: Quality Report to Document Metadata of Migration Surveys

Waleed Mohammed – National statistical office (NSO), Cairo, Egypt

Conducting migration surveys is both complex and expensive, because it requires huge financial resources and large sample to cover the shortage of data on migration, which serve studying recent trends of migration, causes, and its determinants, as well as to clarify the interrelationships between migration and development issues.

In order to maximize the return on such investment it is important that as many secondary users as possible access the data, so documentation metadata for statistical surveys is considered essential when data are to be used by those not familiar with the sources, quality procdures , methods, and details of a database,and this interprete the reasons that make national statistical offices (NSOs) in all over the world are interested to use quality reports to make metadata available for users that contains all information about statistical surveys to acquire users trust in the statistical products.

In this context, Summary quality reports (SQR) is considered one of the ways of communication and building user trust, because it provides them with both quantitative and qualitative information based on European standards. Also, it contains a summary of methods and explains areas where caution in interpretation may be required, structured around the European Dimensions of Quality, this besides, this kind of reports helps users understand statistics and to use it appropriately, On the other hand, SQR is considered very important from the side of NSOs to improve the quality of statistical outputs to update the data ecosystem and make figures transparent and easily accessible to the public.

Based on the above, this paper aims to discuss the Importance of SQR as method for Documentation metadata, and benefits of it for NSOs, and users, by applied the components of SQR on the Egypt Household International Migration Survey 2013 (Egypt-HIMS)" – as a Practical case for how to prepare such type of reports as well as clarify the benefit of SQR in determining challenges' of the Data Ecosystem of migration data in Egypt to update it, and its uses in the documention and archiving qualitative information that related with the statistical surveys in general.

Keywords: Quality management; Statistical awarness; User Satisfaction; Data Ecosystem.

Title of abstract: Managing quality: from CoP and ISO to 'everything'

Douwe Kuurstra – Statistics Netherlands

In 2017 Statistics Netherlands was certified for ISO 27001, in 2018 we were privacy proof and certified for ISO 9001. Now is the time to integrate all types of audits and even more to make a shift from the quality of individual production processes to real quality management.

In the paper I'll will tell you more about the improvements Statistics Netherlands is working on regarding quality; from the demands from individual standards to an integrated approach by an independent audit office.

Background

In 2015 Statistics Netherlands made a roadmap for external certification. ISO 27001, privacy-proof and ISO 9001 were the goals to be achieved. The main reason for certification was an external proof to show our commitment with quality of process and product, privacy and security.

The European Statistical Governance Advisory Board emphasizes in their annual report of 2017: 'Trust among users and the public in general is of the highest importance'. The peers wrote in their report: 'Statistics Netherlands should ... consider also, in the context of external communication, the adoption of recognized industry standard quality assurance/certification schemes.'

After reaching our goals in 2018 we are up for the next level: integrating all types of audits, establishing a more professional audit bureau (with more emphasis on independent auditors) and we changed our focus from individual (statistical) production processes to 'how do all managers know that quality is delivered'. From focusing on individual statistics and standards and we are moving to 'quality for the institution as a whole'.

Keywords: Quality management, auditing, certification, trust.

Title of abstract: A project for sub-municipal register-based statistics in the Italian permanent census: some results of an experimental study

Giancarlo Carbonetti, Marco Di Zio, Fabio Lipizzi, Davide Fardelli, Giampaolo De Matteis, Raffaele Ferrara, Stefano Daddi – ISTAT, Rome, Italy

Since 2018, ISTAT has been conducting the permanent census of the population and housing. The traditional census has been replaced by a combined census based on a system of registers supported by sample surveys. Every year, counts at municipal level are disseminated according to the Base Register of Individuals (BRI), the Base Register of Places (BRP) and a Population Coverage Survey (PCS). BRI contains information on some demographic variables such as gender, place and date of birth,

citizenship, place of residence, derived by administrative data. BRP contains addresses, enumeration areas (EAs) and if possible, geographical coordinates. Both registers suffer from imperfections as missing inclusion of units in the population, inclusion of non-population units, undetected duplicate units and incorrect and missing localization variables. The PCS is designed to estimate under and over coverage rates of the BRI at municipality level. Estimates of population counts are obtained by using BRI corrected with individual weights measuring over and under coverage.

Stakeholders require population counts with increasingly demand for more granular statistics. European Regulations asks for population geocoded to a 1km^2 grid, or municipalities asks for enumeration areas. Integration between BRI, corrected for coverage errors, and BRP allows to locate units on the territory and to produce population counts for sub-municipal domains.

Such integration may fall giving rise to units without an enumeration area. This is mainly due to the quality of address information present in administrative sources, problems relating to the identification and classification of addresses and to linkage errors between the BRI and BRP.

ISTAT is working to improve the geocoding and linkage procedures; for non-geocoded units, different deterministic and probabilistic approaches have been studied for the assignment of the AE according to an imputation process. After an allocation on a geographical data, the geocoding is recovered using administrative data; finally, probabilistic solutions are proposed.

An experimental phase has been defined with the aim to evaluate both the robustness of the different allocation methods and to compare them each other. The tests are conducted on the data referring to a set of municipalities chosen according to: demographic size; levels and quality of geocoding; structural deficiencies of addresses information. Some quality indicators are proposed to evaluate the results. At last, a treatment strategy of non-geocoded units will be defined and implemented in the process of producing sub-municipal census counts.

In general, for register-based census it is essential to implement methodological solutions to assure the quality of the geocoding process. This will have a favourable impact on the accuracy of the census results and on coherence between data related to different statistical universe (individuals, households, dwellings) and referring to the same domain.

Keywords: census; registers; geo-coding; enumeration area; missing data.

Title of abstract: Confidence level and margin of error for non-probability samples: towards a methodological approach

Jeanfrank Teodoro Dantas Sartori, Fabiana Gondim Mariutti – Universidade Positivo, Curitiba, Brazil

Samples gathered without randomization are historically a challenge for both academic and applied management, also being the source of several misunderstandings and misuse. Due to its non-probability nature, classically it is not possible for this type of data to produce inferences to the population as a whole nor to properly define a confidence level and margin of error for the generated outcomes. Some authors and practitioners may sometimes inappropriately adopt a sort of reverse engineering to try to use the same values of a hypothetical random sampling that would result in the same number of collected data, deriving analyses that are not statistically valid. Therefore, this current ongoing research addresses those problems by identifying and proposing a new methodological approach for solving part of this issue, more specifically those cases with categorical data with a finite and known number of options that are analyzed for the frequency of the occurrence of each of those analytical possibilities, provided the gathering sequence is also available. It can also be applied to numeric data, both discrete and continuous if its range is finite. In this view, it is a common scenario for surveys, also known as opinion, attitude, and perception research among practitioners and scientific researchers, with some preponderance in market applications. This proposes approach identifies potential cut points for the sampling either by suggesting stopping further data-gathering or by indicating the possibility of discarding the remaining dataset. Regarding those specific cut points, mathematical properties allow the precise determination of both the confidence level and the margin of error, consequently incorporating statistical significance that is otherwise missing or misestimated. Moreover, it also provides means for the determination of the exact range within which is the frequency for the variable is among the population as a whole (census). This present and prospective development has the potential to contribute to inferential, precise, and statistically grounded use of non-probability sampled data. Even eventually without a large confidence level depending on the sample size, it can provide relevant insights and inferences where probability sample and census data are not available, retrieved, viable or accessible. For that reason, it may provide administrative and quality-centered benefits and strategical-oriented support for both theory and practice, towards the advance of science (e.g., sampling frame operationalization and bold research questions) and business (e.g., decision-making improvement and successful resourcing) among companies, government, and institutions. This new trend may create a ground-breaking research agenda and evolve the existing status quo while proposing further development and validation, especially for extending it to non-finite numerical data applications.

Keywords: Non-probability sample; data collection; data analysis; confidence level; margin of error.

Title of abstract: From indicators to a synthetic index: quality issues in the construction of a small area socioeconomic index

Cristina Rovira, Josep Anton Sánchez, Daniel Ibáñez, Mireia Farré – Statistical Institute of Catalonia, Barcelona, Spain

This paper focuses on the quality impact of the decisions taken in constructing a synthetic socioeconomic index. It also discusses the methodological issues and the solutions adopted in the process of aggregating, weighing, normalizing, and defining a metric for this index. Alternative methodological and metric options are compared. Their impact on the quality of the results is assessed in relation to the goal of accurately representing spatial segregation.

This process is carried out to define and calculate a synthetic socioeconomic index for small areas based on information from administrative registers. The results are presented for Catalonia, but the method can be employed in other territories or countries, adapting it to their own administrative sources.

The small area socioeconomic index (SASI) has been structured in four domains: employment, education level, immigration, and income. A principal component analysis has been performed with an initial set of thirteen indicators; six of them have been selected and combined to define and calculate the synthetic socioeconomic index.

The source of the information is administrative registers so that the index can be computed for small areas (municipalities, neighborhoods, census tracts, 1 km2 grid) and updated annually. For each territory, a value is calculated and expressed as a benchmark Catalonia=100. The definition of the benchmark is unique for all territories, which provides it with broader comparative power.

The creation of the index has been completed with the definition of a division of Catalonia into 853 territorial units of similar population dimensions (10,000 inhabitants) based on socioeconomic criteria. These units include medium-sized municipalities, a grouping of small municipalities and neighborhoods of big towns. They can offer accurate information on existing socio-economic territorial differences or spatial segregation.

Keywords: synthetic index; aggregating, weighting, and normalizing data; small area; spatial segregation.

Title of abstract: Covid-19 vaccination rates: comparison of administrative and survey data

Darja Lavtar, Marjana Vrh, Maja Mrzel, Maruša Rehberger, Nejc Berzelak, Metka Zaletel – National Institute of Public Health, Ljubljana, Slovenia

During the last two years the COVID-19 pandemic has marked the life of the people worldwide. Understanding COVID-19-related human behaviour enables the identification of at-risk target groups and contributes to finding solutions that encourage better adherence to protective behaviour and vaccination recommendations.

In Slovenia, the official data on COVID-19 vaccine uptake, collected in the National electronic registry of vaccinated individuals and adverse events following vaccinations (eRCO), are daily available for the population and some subgroups (e.g. by gender and age groups). The COVID-19 vaccination rate in Slovenia is one of the lowest in the European Union, consequently there are many persons not vaccinated. In order to understand people's behaviour in the pandemic and define relevant measures, additional information is needed from other data sources.

In December 2020, the Slovenian National Institute of Public Health (NIJZ), in the Slovenian statistical system holding a role of the Other National Authority (ONA) for health and health care statistics, launched the Behavioural Insight Survey on COVID-19, a National Survey on the Impact of the Pandemic on Life (SI-PANDA 2020/2021). The survey consisted of nineteen waves that were conducted once every two weeks in the first half of 2021 and once a month in the second half of 2021. Each wave of the online survey includes a sample of about 1,000 respondents aged 18 to 74 years. The survey data on the vaccine uptake enable comparisons by gender and age groups (like administrative data), as well as by other socio-demographic and explanatory variables. Providing relevant, accurate and timeliness data on the behaviour of both vaccinated as well as non-vaccinated population can be crucial for data-based decision making. Since administrative data have limited explanatory power, the use of survey data can complement with information on the other aspects of pandemic behaviour (e.g. the educational level, the economic activity, the reasons for not taking the COVID-19 vaccine, etc.). The survey data on vaccine uptake is also important as an explanatory variable, enabling the understanding of the adult population behaviour in the time of COVID-19 pandemic.

The paper evaluates the accuracy of the Slovenian vaccine uptake data collected in eleven survey waves of SI-PANDA by comparing them to the administrative data from the eRCO register. Comparisons are performed at the level of the entire target population as well as by gender, age groups and vaccine doses (first, second, additional) to identify potential differences in the accuracy of survey estimates by subgroups.

Keywords: COVID-19 vaccination rate; administrative data; web survey data; timeliness; accuracy.

Title of abstract: Using data visualization for achieving the quality of statistical products (Economic census Application)

Mahmoud Mohamed ElSarawy, Reem Ismail Mohamed ElSybaey – Central Agency for Public Mobilization and Statistics (CAPMAS), Cairo, Egypt

The data visualization tools promote the best use of evidence in decision making, transparency of data and information and people comprehend data better through pictures than by reading numbers in columns and rows. The main goal of data visualization is to communicate information clearly and effectively through graphical means, so all users have to do is authenticate and click a button to create infographic or data visualization and shared with their friends online in social media. It offers the ability to create different views of your data and change them as your needs. The full positivity of the quality of statistical work is achieved by achieving the maximum benefit for the data user and decision makers, through an accurate and clear reading of the data, which is achieved by visualizing the data.

This paper will study the importance of using data visualization in our statistical world due to the large amount of data and needs to be analyzed quickly so it will get useful impact on data users and decision makers. This will be applied to the latest Egyptian economic census data that was recently implemented in Egypt using a new technological system which helped to collect data electronically and display the results in a faster time and to read the data properly, with high quality, in a comprehensive and easy-to-understand manner for the decision maker. Moreover, analyze the relation between data visualization and future statistical challenges like big data and statistical awareness.

Keywords: Data Quality, Infographic, Decision makers.

Title of abstract: How to disseminate quality information in a user-friendly way? Revised methodological guidelines in development

Petra Fekete-Nagy, Erika Földesi, Katalin Kovács, Katalin Szép – Hungarian Central Statistical Office, Budapest, Hungary

The quality and availability of statistics has become again a particularly accentual central issue since the pandemic. Users and their needs have multiplied, putting considerable pressure on National Statistical Institutes. The situation has also led the Hungarian Central Statistical Office to rethink its usual practices. We have developed a comprehensive guideline that specifies where metadata should be provided, how and what should appear on the website or on downloaded tables.

We developed a single set of guidelines for different types of data and indicators that are both professional and user-friendly. For this purpose, the different quality information were classified into 7 main categories (relevance, accuracy, timeliness, comparability, accessibility, methodology, other), and then we determined the level

at which the metainformation for the given data should be reported. We have provided details on which ones should appear right next to the data, which ones should be more distant, one or more clicks away. We have also defined their exact location and appearance.

The guidelines are the result of several sub-processes:

We obtained information from several sources to assess user needs. Firstly, we examined the content of the questions received from the users through the Contact Centre of the HCSO, about information content that is missing or not clear for our users.

In addition to this, we conducted a minor research within a specific target group: we asked university students to comment on the topic based on a questionnaire we compiled. Furthermore, we relied on the expertise and experience of our colleagues in our office who are dealing with information policy.

In order to develop requirements, we analysed the reporting practices of other NSIs and Eurostat, processed the literature and consulted with subject-matter domain experts on how to interpret the individual quality information in relation to the different key indicators.

Our proposal was validated again with users to develop the final format. The introduction procedure is continuous.

The topic generated further professional discussions and forums in our office, for example, about the information related to the accuracy of the data and their communication. To what extent do we know the accuracy of our published data, what do we do to find it out? Are we helping the user or are we just making them uncertain if and how we inform them on the accuracy of our data?

Keywords: online dissemination, quality information, user needs, metainformation.