

5. The institutional transformation of NSOs – New methods, services and roles (Moderator Mariana Kotzeva)

This session will explore the new ways that NSOs are producing estimates, disseminating data and providing access to their users. The discussion will focus on whether the existing frameworks are properly leveraging these new data sources and methods and if not, what steps need to be taken to ensure these new methods, services and roles are incorporated into our national and international infrastructures.

- b) New services including microdata access and linking. ([Eurostat and Australia to provide discussion document from Global/National perspectives](#))

There is a growing impact of Multi National Enterprise (MNE) groups on the exchange of goods and services at the European and global level. Globalisation represents one of the most prominent challenges for (national) statistical offices to produce economic statistics at the national level. The exchange of data between NSOs is necessary to cope with these challenges.

Within the European Union, Eurostat, in the context of the European Statistical System (ESS), has engaged in enabling and facilitating the exchange of microdata for the purpose of enhancing the quality of business and trade statistics and increasing the efficiency of the overall statistical system. For this purpose, Eurostat is working on legal enablers but is also supporting the establishment of governance structures and the implementation of IT infrastructures for data exchange within the ESS. The most prominent examples are the Regulation on European Business Statistics (FRIBS) and the exchange of micro-data for the EuroGroups Register.

The integration of new data sources for the regular production of official statistics requires innovation at multiple levels, including new processing paradigms, computation methods, data access and governance models, staff skills, etc. The term "Trusted Smart Statistics" was coined by Eurostat to indicate a comprehensive framework to evolve official statistics towards adoption of new data sources along with traditional ones. Eurostat is currently elaborating principles, which might affect the way how data will be used by statistical offices in the future.

Based on the forthcoming implementation of the Regulation on European Business Statistics (FRIBS) starting from 2021, two micro-data exchange processes in the ESS will be legislated:

- a) The micro-data exchange on intra-EU trade in goods statistics between national statistical authorities

From January 2022, FRIBS will make mandatory the exchange of micro-data on intra-EU exports (MDE) between the EU Member States. This will create an additional data source for the compilation of the intra-EU imports of goods. With the help of the MDE, the receiving (importing) Member State may collect less data (or no data at all) from the domestic importing

companies to reduce the administrative burden, while having the flexibility to develop its compilation methods and ensure sound level of quality.

Micro-data will be exchanged through a centralised hub hosted by Eurostat, the MDE Hub. Each Member State will send to the MDE Hub micro-data on its exports to other Member States. The whole process of data transmission from sending (exporting) Member States to receiving (importing) Member States via the MDE Hub will comply with high security standards. The sensitivity of micro-data will also make necessary the establishment of very strict rules as regards the data retention by the MDE Hub and use by the receiving Member States.

b) The microdata exchange with the EuroGroups Register.

The increasingly global activities and structures of enterprises challenge the integration, coherence and comparison of business- and macroeconomic statistics across Member States and across domains.

The EuroGroups Register (EGR) is the basic statistical ESS infrastructure for the coordination and storage of basic information about MNEs. Using the microdata input from the Member States, the EGR links and contains data on the structure of MNEs resident in the EU and their constituent legal units and the underlying economic variables.

The EGR provides consistent and timely information on MNEs to the ESS, serves as data source and survey frame for MNEs (their structures and economic variables) when producing national and European statistics, enables to enrich the National Statistical Business Registers (NSBRs) with information on MNEs consistently produced across the ESS and provides consistent data of extra-EU units belonging to European MNEs.

The EGR provides insights in the structure and behavior of internationally operating enterprise groups that cannot – or only with huge efforts - be obtained on the national level solely. Data for more than 121,000 MNEs active in Europe are available to users in the ESS.

More microdata exchanges also enable more microdata linking of the various data sets of European business statistics. This means that more relevant indicators can be produced with relatively low costs.

In the following we describe a subset of Trusted Smart Statistics principles that might be relevant in the context of exchanging and using data in an international context.

Modular methodological frameworks

Data are often a by-product of technological processes that are subject to change following the natural evolution of technologies and/or usage habits. This introduces temporal changes in the source data. Moreover, for some classes of data sources, certain detailed aspects of the technology are not completely standardised and may vary across countries, which introduces changes to the detailed formats and structures of data generated by different sources and/or across different countries. Heterogeneity and non-stationarity of input data details poses additional challenges for the development of processing methodologies in terms of evolvability

(to cope effectively with changes in time) and portability/interoperability (to address heterogeneity across countries and individual sources). The key to address such challenges is to develop highly modular methodological frameworks, where each module can be evolved or replaced without requiring changes to the rest of the processing workflow.

Using data without sharing data / pushing computation out

In many cases, new data sources are collected and held by organisations that don't belong to the statistical system, including private companies. One obvious way to process such data for official statistics is to "pull data in", move data inside the statistical system, and to process it internally. This approach might not be the most convenient in all settings, e.g. when data are business confidential, highly privacy-sensitive, or when the volume and rate of input data is much larger than the required output information. In fact, the focus of the statistical system should be on the extraction of the desired output information, i.e. the final statistical indicators. The acquisition of input data should be seen merely as an ancillary (but not always necessary) task. Executing certain (pre-)processing functions at the source premises, and then retrieving the intermediate data instead of the raw input data represents one step away from the "pulling data in" towards "pushing computation out" paradigm.

In more advanced scenarios, the desired output statistics is obtained by fusing input data held by different entities. In this case, the adoption of modern techniques from the field of Secure Private Computation (e.g. Secure Multi-Party Computation (SMPC)) enables statistical offices to extract the desired output information, representing intermediate data or even final statistics, without having direct access to the raw input data. Such technologies appear to be now mature for adoption in production processes. In the same way as extracting information instead of internalising data from third parties, information could be extracting from statistical data repositories held by different statistical offices. In case of confidential data secure private computation methods can be used to reduce obstacles to data access and at the same time prevent risks of data exchange and concentration. The combination of SMPC with advanced automated methods for dynamic Statistical Disclosure Control (SDC) can help to protect statistical confidentiality in multi-source contexts.

Eurostat, together with the ESS, is currently working on developing methodological frameworks for selected types of big data, e.g. mobile network data, and on uses cases for pushing computation out and application of secure private computation, e.g. in the context of financial transaction data with central banks.