Quality assurance in micro-data exchange – The international trade in goods statistics as concrete example

Anne BERTHOMIEU, Eurostat, anne.berthomieu-cristallo@ec.europa.eu

Abstract

The objective pursued when exchanging micro-data is to benefit from an additional data source with no extra burden for the providers of the statistical information (e.g. enterprises). Such a data exchange requires a performant, robust and secure transmission system, embedding comprehensive guidelines and binding rules for both the sender and the receiver of the information. Rules binding the receiver are mainly targeting the respect of the data confidentiality while the sender is mainly bound to quality-related targets covering the data completeness, accuracy and punctuality.

The quality assurance is built on five pillars: a requirement system composed of binding rules on who should do what, how and when; a guidance system aiming at supporting the stakeholders and promoting good practices; a reporting system composed of all types of metadata to be attached to the data flow or to be provided as an additional component; a monitoring system gathering all information on the data flows and the quality of the exchanged data; and an assessment system producing compliance reports pointing out where actions are needed. Such quality framework is on the way to be implemented for the exchange of micro-data relating to intra-EU exports of goods.

**Keywords:** quality assurance, requirements, monitoring, reporting, quality assessment

1. Exchange of micro-data on intra-EU trade in goods

1.1 Introduction to intra-EU trade in goods statistics

Intra-EU trade trade in goods statistics measure the value and quantity of goods traded between the EU Member States. ‘Goods’ means all movable property including electricity. The Combined Nomenclature classifies the goods for which statistics are collected on the basis of around 9 500 product codes.

The current data collection system for compiling intra-EU trade statistics (Intrastat) can be characterized as follows:

* data are directly collected from traders;
* it is closely interlinked with the VAT system relating to intra-EU trade to ensure the completeness and quality of the statistical data; and
* a system of thresholds is established to simplify data provision and reduce the overall burden on traders, particularly small ones.

1.2 Reducing the response burden while maintaining the data quality

As the burden associated with Intrastat was considered to be quite high, in November 2011, the European Council called upon the European Statistical System (ESS) to take effective measures ensuring a substantial reduction of the response burden by redeveloping Intrastat, while maintaining at the same time, a sound level of quality.

In response to this call, the ESS adopted an innovative statistical approach consisting in exchanging micro-data on intra-EU exports between the EU Member States allowing them to use those mirror data for the compiling of their own intra-EU imports statistics. Such an approach follows the principle that data collected and available within the ESS need not to be collected more than once. Thus, each trade transaction collected in one Member State may serve as a data source for two Member States: first, for compiling the intra-EU exports of the exporting Member State and, second, for compiling and/or verifying the intra-EU imports of the partner Member State. As a consequence, the benefit of the approach is two-fold:

* reduce the administrative burden on Intrastat reporters on the imports side; and at the same time
* guarantee that the statistical information provided is fit for the purpose and of high quality.

1.3 Technical system for micro-data exchange (MDE)

A regular and voluminous exchange of micro-data on intra-EU trade in goods between Member States during the data production cycle needs a new, sophisticated and high-performance IT solution due to its security and architectural dimensions.

A pilot project, SIMSTAT, was carried out in order to prove both the technical feasibility of the exchange of micro-data on intra-EU trade in goods and the feasibility of the use of the exchanged data as an additional data source for compiling intra-EU imports statistics. An exchange of micro-data took place through a new IT system which was delivered in the 1st quarter of 2015 and tested through pilot exchange by 20 Member States during the 2nd and 3rd quarters of 2015.

Various technical options were investigated, the merits and drawbacks of each of them analysed, the risks and benefits evaluated and the implementation costs assessed. On that basis, the creation of a centralised Hub-based system was recommended as the most advantageous option. In this system all the Member States should submit data according to standard processes, formats and technologies from their systems to the data Hub; the data Hub should split the received files by partner Member States and forward data to the relevant partners.

Figure 1. High level architecture of the system



Another technical outcome of the SIMSTAT pilot exercise was that exchanging incremental files instead of full datasets was more efficient in terms of IT resources due to the high volume of data. However this implies that the data Hub should embark all the necessary functionalities to automatically detect and correctly manage anomalies in the data files transmission sequence or the data files content.

In addition to the centralised data Hub solution, it was also recommended to use CCN network, owned by Directorate General TAXUD, as the transmission channel. This network is currently being used by the Member States for VAT and VIES data exchange between Customs and tax administrations.

Data security plays a crucial role in the micro-data exchange. Being at trader level, the micro-data are, by definition, confidential. They have to be treated and transmitted securely in compliance with EU regulations. The data Hub will be hosted in a secure infrastructure segregated from the standard infrastructure of the European Commission, establishing secure connections with IT platforms and providing services allowing accredited users to interact with confidential data that they have permission to access. All data transmitted over the network will be encrypted from end to end.

1.4 What are the MDE stakeholders and respective roles?

The MDE stakeholders are the National Statistical Authorities of the EU Member States and Eurostat. Member States are both data compilers and data users:

* The exporting Member States are liable to compile the statistical data and to provide the metadata necessary to the MDE Hub to ensure the integrity of the data exchange and to the partner (importing) Member States to interpret the data and assess their quality; and
* The importing Member States are liable to store and use the received data according to the established rules, in particular respecting their confidentiality.

Eurostat is the system provider whose role consists in setting up and maintaining the communication channels and the system (i.e. MDE Hub) which will process the incoming data files and will produce the outgoing data files destined to the receiving Member States. It should also provide a technical and methodological support.

In addition, Eurostat should act as a supervisory body by ensuring a daily monitoring of the data exchange, reporting any issues and ensuring a follow-up of the corrective actions to be possibly taken.

1.4 What more in micro-data?

Micro-data on intra-EU trade in goods relate to information provided at trader level. Beyond the level of details, they also include not only the standard statistical data elements for intra-EU exports but also complementary information necessary to the receiving Member State to build its own intra-EU imports statistics. In addition to the identification of the exported goods and the nature of the transaction, the importing Member State needs to identify which national trader is involved in the transaction and possibly the country from which goods are originating when this country differs from the exporting Member State.

The micro-data exchange also requires a set of operational metadata to be used by the MDE Hub to properly manage the incoming (from the sending Member State to the MDE Hub) and outgoing (from the MDE Hub to the receiving Member State) data, ensuring no rupture in the flow and controlling the consistency of the data files sequence and content.

1. The MDE Quality Framework

The quality assurance in MDE is ensured by a quality framework built on five pillars:

* A requirement system composed of binding rules on who should do what, how and when;
* A guidance system aiming at supporting the stakeholders and promoting good practices;
* A reporting system composed of quality monitoring reports and reference metadata to be attached to the data flow (monitoring metadata) or to be provided as an additional component (structural metadata);
* A monitoring system based on operational information on the data flows and the quality of the exchanged data; and
* An assessment system producing compliance reports and listing the actions possibly needed.

Figure 2. Overview of the MDE Quality Framework

2.1 Requirement system

The requirement system is composed of the rules binding all the stakeholders of the micro-data exchange and covering all the steps of the MDE statistical process from the data collection (i.e. compilation and sending by the sending Member State), treatment (i.e. validation and processing by the MDE Hub) and use (i.e. reception, storage and dissemination by the receiving Member State). Some rules are of legal nature, i.e. laid down in the European legislation (e.g. mandatory versus optional data elements, data transmission deadlines); others correspond simply to recommendations aiming at improving the data quality by promoting good practices at national level (e.g. confidentiality practices, revision policies) or to compilation instructions (e.g. structure of the data files, codification of the information according to the relevant nomenclature).

Some requirements are translated into validation rules to be passed by the data before their delivery to the partner Member States. Validation steps ensure that data are the expected ones (e.g. transmission of all mandatory data elements) and that the minimum quality targets are reached (e.g. intra-dataset consistency). Validation rules need to be linked to error severity degrees and related actions. Errors considered as critical should lead to the rejection of the whole file even if relating only to a minority of records. Non-critical errors could lead to simple warnings to alert the receiving Member States on accuracy or consistency issues. The cases where the MDE Hub should stop the data flow and automatically request the transmission of a revised data file should clearly be identified.

2.2 Guidance system

The guidance system is composed of reference documents whose purpose is to guide the stakeholders and of referent bodies whose purpose is to define the strategy (high-level) and to prepare the concrete implementation steps (expert-level).

The core reference document is the MDE Handbook which seeks at answering the needs of all types of stakeholders. It covers the micro-data preparation and sending by the exporting Member State, the micro-data processing by the MDE Hub and the micro-data reception by the importing Member State. It also includes specific chapters on security and users management, quality assurance, cooperation between national administrations and technical support.

The core referent body is the MDE Task Force which is responsible for the technical preparation and readiness of the micro data exchange among the Member States. Such a referent body is essential not only for the implementation phase but also to ensure that the system keeps being reactive to issues and evolve whenever necessary to fill in the possible gaps, strengthening the functionalities or answering new needs. The MDE Task Force should report to and get guidance from its parent group, the Steering Group on international trade in goods statistics. All Member States are members of this task force which is chaired by Eurostat.

The guidance system is completed by a user support ensured by the system provider, Eurostat, to answer technical and methodological questions.

2.3 Reporting system

The reporting system encompasses reference metadata to be provided by the sending Member States and quality monitoring reports to be prepared by the system supervisor, i.e. Eurostat.

Reference metadata enables the receiving Member State to correctly interpret the data and assess their quality. They can be either attached to the micro-data files or provided as an additional component. They are of two different sorts:

* the structural metadata which gather the descriptions of national compilation methods and practices (e.g. revision policy or confidentiality practices); and
* the monitoring metadata which correspond to quality indicators related to the transmitted data files (e.g. coverage from collected data or revision rates).

Reference metadata should be distinguished from operational metadata (e.g. file number or total values to be compared to the sum of all records) to be used by the MDE Hub to check the consistency of the data files transmission sequence or the data files content.

In addition to metadata, a quality monitoring report should be produced at least once a year. Its purpose is to provide a feedback to all stakeholders on the global functioning of the micro-data exchange.

2.4 Monitoring system

The monitoring system gathers all the tools and procedures in place to assess to which extent the MDE requirements are met. One of the key tools is the data validation module which checks the format and content of the data files and automatically produces an error report every time micro-data are processed by the MDE Hub. Although the objective is to transmit error-free datasets, in the real life, data files are often impacted by issues. Some are minor (format-related ones) and can be easily corrected, others may require adapting the data preparation process or even to contact the provider of the information (i.e. intra-EU trader). To which extent a data file complies with the requirements needs to be monitored.

Other tools and procedures aim at assessing quality targets like the data punctuality (e.g. micro-data files to be delivered within 30 days after the reference month) or accuracy (e.g. micro-data files produced from collected data amounting for at least 95% of the total trade). Such targets are assessed on the basis of monitoring metadata either collected or derived from transmitted data.

Although not related to a specific target, it should be noted that indicators on data revisions are important quality measures. They should then be also monitored, which requires the MDE Hub to store data vintages at least at aggregated level.

Requirements relating to compilation practices (e.g. way to hide confidential data) are monitored on the basis of transmitted data and/or structural metadata.

2.5 Assessment system

The assessment system is based on compliance reports produced by the system supervisor (i.e. Eurostat) and on the feedback provided by the data users (i.e. the receiving Member States).

A compliance report should be prepared at least once a year on the basis of the quality monitoring reports and of any relevant additional information. Its purpose is to list the cases of non-compliance with the MDE requirements as well as to highlight where actions are needed and to ensure a follow-up until issues are solved.

The receiving Member States should also provide a feedback on the data and metadata received. By considering information available at national level, they should assess the data quality and report any issues at least on an annual basis.

1. Conclusions

The micro-data exchange in the international trade in goods area is an ambitious reply to the challenge of reducing the statistical burden on providers of the statistical information (here the intra-EU traders) while maintaining the data quality. It requires strong involvement and commitments of all parties involved (National Statistical Authorities and Eurostat), binding rules and sound guidelines to set up the whole process and handle any statistical or technical issues in the appropriate manner.

Sending Member States commit in compiling and transmitting the data meeting the requirements, Eurostat commits in setting up a robust and performant IT system meeting high requirements in terms of security and data controls and for which failures are not allowed, and receiving Member States commit in using the data according to the established rules.

A close cooperation between the ESS members and a quality framework strengthening the mutual confidence are the keys for the success of such a micro-data exchange.

1. References

Eurostat Statistics Explained – [Intra-EU trade - exchange of micro-data](http://ec.europa.eu/eurostat/statistics-explained/index.php/Intra-EU_trade_-_exchange_of_micro-data)

ESS Vision 2020 project SIMSTAT – [Project-End Report: ESS.VIP SIMSTAT](http://ec.europa.eu/eurostat/documents/6842948/0/SIMSTAT%2BProject%2BEnd%2BReport%2Bv1.2%2B-%2BWeb/aafe9259-c8ce-4598-80d9-25fcf7f04d3b)

ESS Vision 2020: Building the Future of European Statistics – <https://www.youtube.com/watch?v=i5VRp6mrAjU>

Eurostat Statistics by Theme – [International trade in goods](http://ec.europa.eu/eurostat/web/international-trade-in-goods/overview)