Production of International Trade Statistics in Eurostat

Ján Plánovský 1

Summary

Paper describes production process of international trade statistics in Eurostat. The detailed data on trade in goods between 27 Member States of the European Union (Intrastat) and the statistics on the EU trade with non-member countries (Extrastat) are collected on a monthly basis from the European trade operators and transmitted to Eurostat via relevant national statistical authorities. Eurostat further compiles the official statistics for the EU and euro area from the data provided by the Member States.

Production of trade statistics is characterised by huge volume of data (around 9500 product codes, more than 200 partner countries, multiplied by 2 flows and dozens of indicators), application of confidentiality on certain flows of goods requiring special data treatment in highly protected environment, and uniqueness of trade statistics processing in a way that the whole production chain starting from data reception up to dissemination and user support is managed by one production unit imposes wide and complex demands on staff expertise.

Ongoing improvement and development of customised IT tools, automation of production processes, harmonisation of methods and validation rules between Eurostat and Member States, reduction of statistical burden, changes in the customs clearance data collection and transmission systems and preparation of common quality standards represent top priorities for present and near future of trade statistics.

¹ Ján Plánovský is Head of Unit "International Trade – Production" at Business Statistics Directorate of Eurostat, European Commission, Luxembourg. E-mail: jan.planovsky@ec.europa.eu

Introduction

1.1 Eurostat's mission

The mission of Eurostat, the Statistical Office of the European Communities, is to provide the European Union with high-quality statistical information at European level that enables comparisons between countries and regions. Established in 1953, the tasks of Eurostat have evolved and broadened over the years. With the development of Community policies, such as the EMU, Eurostat adapted its role to respond to the need for new statistics. Today, Eurostat is one of the Directorates-General of the European Commission, based in Luxembourg. In order to fulfil its mission, it is committed to

- providing the Commission with the **high-quality statistical service** needed to develop, implement and evaluate policies;
- developing a partnership with the corresponding statistical services of the European Central Bank (ECB);
- producing, with the assistance of the European Statistical System (ESS), reliable, comparable and relevant statistics covering the EU's areas of competence;
- disseminating Community statistics to the general public, businesses and decision makers, as part of its role as a public service provider, free of charge;
- **supporting non-EU countries**, particularly candidate countries, which wish to develop their statistical systems within the framework of the EU's external relations with those countries.

Eurostat operates in full accordance with the "European Statistics Code of Practice" [3] . Modern societies do not function properly without a solid basis of reliable and objective statistics. On one hand, decision-makers at EU level, in Member States, in local government and in business need statistics to make those decisions. On the other hand, the public and media need statistics for an accurate picture of contemporary society and to evaluate the performance of politicians and others. Of course, national statistics are still important for national purposes in Member States whereas EU statistics are essential for decisions and evaluation at European level. Specific role of Eurostat in the area of international trade statistics include:

- Methodology and classification unit is responsible for providing the European Union with the
 legislative and methodological framework for the collection, processing and dissemination of
 quality foreign trade statistics. Preparing, adapting and updating the Intrastat and Extrastat
 legislation in the area of foreign trade statistics taking account of user needs, the economic
 and administrative environment and the reporting burden on respondents is the core activity.
- Production unit is responsible for collection, validation and further processing of detailed intra
 and extra EU trade, electronic dissemination to the public, the Commission, the European
 institutions and international organisations through the Comext on-line system which is free of
 charge. The quality assessment of published data, through quality reports, analysis of
 asymmetries, outlier detection and specific studies is integral part of this work.

1.2 Basics of international trade statistics

The international trade statistics of the EU cover both extra- and intra-EU trade of goods, with "goods" meaning all movable property including electric current:

- Intra-EU trade statistics cover the trading of goods between 27 Member States ("Intrastat");
 this system is closely linked to the Value Added Tax system.
- Extra-EU trade statistics cover the trading of goods between Member States and non-member countries ("Extrastat"); this system is closely linked to customs rules and procedures.

The need for statistics on trade in goods is related to the role international trade plays in the economy and, as such, must be measured reliably and the relevant statistical data should be comparable and widely disseminated.

The statistical information is mainly provided by the trade operators on the basis of Intrastat declarations (Intrastat) or Customs declarations (Extrastat). Data are collected by the competent national authorities of the Member States which then compile the detailed monthly statistics according to a harmonised methodology established by EU regulations and transmit the results to the Commission (Eurostat).

The system for collecting statistics on the trade in goods has undergone major changes starting with the introduction of the Combined Nomenclature for the coding of goods (CN) and the Single Administrative Document (SAD as customs declaration, used also as source of statistical data) in 1988.

Intrastat is the result of removal of customs formalities between Member States within the Single Market from 1st January 1993. A new system had to be established for collection of data on intra-Community trade. Whereas Extrastat was and still is largely built on administrative data collected under customs procedures, and thus assures a high level of control and data verification, the introduction of Intrastat meant a methodological break as data are directly collected from the economic operators.

A lot of effort has been made and continues to assure the quality of data in Intrastat, while at the same time rationalising the statistical system and reducing the burden on data providers.

1. Trade Statistics Data Processing

This section describes in more detail trade statistics data processing environment in Eurostat. Extra-EU trade data originate from the Customs Authorities of the Member States (MS) which use the Single Administrative Document (SAD) as the documentary source for customs declarations on the trade in goods with non-EU countries. Up to 38 fields of the SAD - over a total of 54 - are further used by the Customs and/or by the National Statistical Institute (NSI) in order to compile extra-EU trade statistics Intra-EU data originate from Intrastat declarations submitted by trade operators which are structured like SADs, with the difference that certain operators falling below the reporting threshold are excluded from this collection.

2.1 Data reception, validation

The NSIs of 27 Member States collect and process trade data according to the harmonised rules. At least once a month each NSI is obliged to transmit to Eurostat the data aggregated at CN8 product code level according to the predefined publication calendar. Typical monthly sending for one MS includes 2 data sets:

- 1. Intrastat arrivals and dispatches (26 partner countries)
- 2. Extrastat imports and exports (all non-EU partner countries²)

One monthly data set can contain the records for several consecutive reference periods. The frequent practice of MSs is that the data from January to the last available reference month are sent in every monthly transmission. The legal deadline for sending the detailed data is 70 days for Intrastat and 42 days for Extrastat after the end of reference month.

Table 1: Volume of data processed for reference year 2008

	Intrastat	Extrastat	TOTAL
# monthly periods including revisions - # first deliveries:	1594 12*27=324	1889 12*27=324	3483 648
# Gbytes	32.0	21.5	53.5
# records (millions)	275	185	460

As the Table 1 indicates, the first data deliveries represent 324 (12 months times 27 MSs) reference periods, both for Instrastat and Extrastat. The actual amount of transmitted data represented 3483 individual reference periods, which means that each month of data has been revised approximately 4 times per year. Certain MSs routinely transmit the data for current reference period plus data for 15 preceding months. So the volume of data to be transferred and validation might multiply 10-fold. Total amount of 53.5 GB of raw data corresponds to 460 million individual records. Simple division suggests that the average number of records per one reference period sent by one MS is around 132,000 records.

According to Anne Berthomieu [1], over the EU about 70 million SADs were collected in order to cover the trade relating to 110 million items for the year 2007. Around 150 millions of supplementary items were reported through simplified declarations.

_

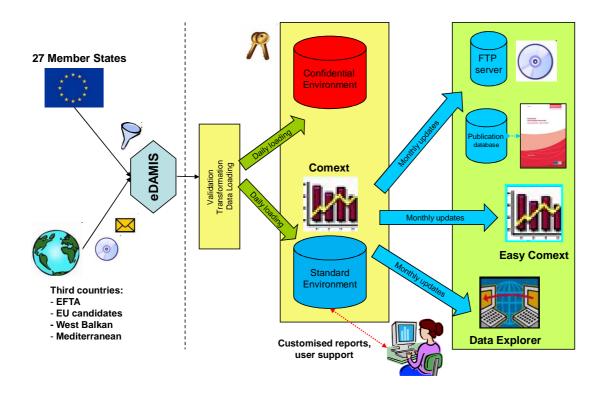
² There are 259 countries and territories listed in Eurostat Comext database, including those not existing any more (e.g. Yugoslavia).

Schematic trade data flow is illustrated in Figure 1 below. The data of EU members are received solely electronically in GESMES format via Eurostat's single entry point, the so called eDAMIS web-based application. Other trade data sources are the third countries which also regularly transmit data for processing in Eurostat. They include 4 EFTA states, the candidate and potential candidate countries for EU membership and countries participating in various regional cooperation programmes (e.g. West Balkan, MEDSTAT, TACIS). This data might be received by another technical means like CDs/DVDs or e-mails depending on local IT infrastructure. Logically, only extra-EU (Extrastat) statistics are relevant for non-EU countries.

Once the data flow crosses the virtual vertical dotted line they are treated under the responsibility of the trade statistics domain in Eurostat. Because of its volume, the international trade statistics data can only be validated through automatic procedures. In Eurostat, all the data files transmitted by the EU Member States are automatically processed by a dedicated production system. The subsequent yellow box fed with data by eDAMIS represents the Data Production System (DPS) which validates the plausibility of CN codes and country codes and performs several logical tests for variables which are mutually dependent. The system also points out missing, invalid and inconsistent entries for all provided statistical variables and detects outliers in trade indicators - statistical value and net mass. Some errors are considered as "fatal", which means that the automatic data processing stops and a manual intervention is then necessary for investigation and decision. Fatal errors are mainly referring to missing, invalid or inconsistent entries impacting key variables, i.e. variables considered as essential for the users and/or on the basis of which other variables are validated (e.g. key variables are the type of trade flow - import or export - or the type of data - collected or estimated by the NSI). In most cases of fatal errors, a data revision is requested from the Member State concerned. Only in exceptional cases, the erroneous entry is manually corrected or the interval manually extended. "Non-fatal" errors correspond to missing, invalid or inconsistent entries impacting non key variables. Such errors are automatically corrected using default values and listed in a report indicating the

Figure 1: Trade Statistics Production in Eurostat

original and corrective values.



After the validation steps are completed, the DPS transforms data into appropriate structure ready to be loaded into the database. It is important to separate confidential records which are supposed to be loaded into special confidential environment, the so called BRI (internal reference database). All other records are stored in standard environment, so called BRE (external reference database). The BRE represents the core of main production and dissemination system named Comext (from commerce extérieur) implemented in Oracle. Both BRI and BRE are updated on daily basis according to the first or revised data reception from MSs.

2.2 Data dissemination, user support

Comext is the Eurostat reference database for external trade. It provides access not only to both recent and historical data from the EU Member States but also to statistics of a significant number of third countries. Connection to Comext is officially granted to the internal users of the Commission and any other European institution or Community body. The database is also accessible to the administrations of Member States belonging to the European Statistical System and to those Candidate Countries providing statistics to Eurostat.

Comext database is locked for loading once a month for the 3-days period to update the dissemination environment (see the green box in Figure 1) with new reference period in order to prepare monthly Press Release [8]. While Comext is an on-line analytical and dissemination tool requiring special training, Easy Comext [5] is its lighter version destined for general public dissemination. It is updated once a month according to a predefined data release calendar.

Before inputting the data into Easy Comext additional visual control of trade with all partner counties is performed. There are 4 high-level aggregate tables automatically produced for each of 27 MSs resulting in 108 graphics where the time evolution of trade is visible. For example, in case of Italy the following combinations are generated:

- 1. Total imports of Italy from every partner country of the world, monthly series for the last 30 periods, expressed in quantity (ton);
- 2. Total imports of Italy from every partner country of the world, monthly series for the last 30 periods, expressed in value (1000 of euro);
- 3. Total exports of Italy to every partner country of the world, monthly series for the last 30 periods, expressed in quantity (ton);
- 4. Total exports of Italy to every partner country of the world, monthly series for the last 30 periods, expressed in value (1000 of euro).

Figure 2 below illustrates such graphical result for the case 4. Each line corresponds to one trading partner with the exception of top six lines which represent selected geographic or economic zones. Two lines further down (blue and red) represent the main recipients of Italian exports - Germany and France. Seasonal effect showing the drop in trade activity during August summer holidays is clearly visible. Besides that, the overall decline in exports since October 2008 could be attributed to global economic crisis. Any major errors or deviations are identified in similar way and further investigated before data are loaded into main dissemination environment – Easy Comext.

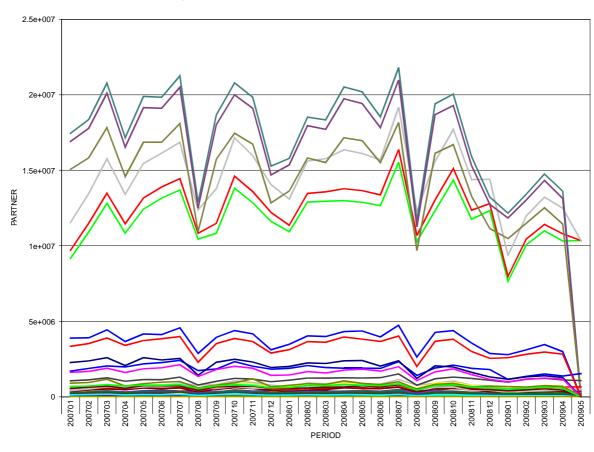
Other dissemination channels are:

- **Comext DVD-ROM** with Comext stand-alone dissemination interface; contains selection of external trade statistics for the most recent months, data broken down by Combined Nomenclature product and intra- and extra-EU partner country. It also contains PRODCOM database (data on production, external trade and domestic consumption of more than 5000 products). The Comext DVD runs under Windows and offers a wide variety of functions. It is intended mainly for libraries and subscribers which don't have access to high-speed internet.

There are 11 monthly issues of DVD (except July) including monthly data for current and previous year and 3 supplements:

- 1. Monthly data since 1998
- 2. Annual data since 1988
- 3. Trade by mode of transport.

Figure 2: Visual validation of Italian exports by partner on aggregated level



DECLARANT: Italy, PRODUCT: TOTAL, FLOW: 2, STAT_REGIME: all, Indicator: VALUE_1000EURO

- **Bulk download** facility - FTP server containing packed raw data in flat file format is dedicated for bulk downloads of data, used mainly by registered commercial redistributors.

Publication database represents an internal production tool facilitating the preparation of regular short-term indicators and long-term series, such as:

- Euro-indicators News Release [8] published monthly
- Monthly Bulletin on External and intra-European Union trade [7]
- Statistical Yearbook External and intra-European trade [6]
- Pocketbook [10] published bi-annually
- Panorama [9] published once in 4-5 years
- Statistics in Focus –3-4 issues per year on dedicated topic.

All these publications were designed to be printed, so they are created in PDF format. Limited number of printed copies is published and distributed via Publication's Office of the European Communities.

Besides that, both short-term and long-term statistics are disseminated via Eurostat's website by TGM (Tables, Graphs and Maps) or Data Explorer interfaces.

Comext also serves as a data source for calculation of the seasonally adjusted series and unit value and volume indices on EU trade. While seasonal adjustment is processed by the software DEMETRA that carries out corrections for working days as well as launching a method of seasonal adjustment known as TRAMO/SEATS, indices are calculated by another internal special application. The resulting series and indices are loaded into publication database. They are finally disseminated via TGM interface to general public.

An integral part of dissemination service is support of users. Eurostat has recently developed and implemented the tool for harmonised user support — XT-NET/Assist (see Figure 3) to facilitate a fully integrated environment which would cope with user requests within the whole ESS. The web-based application enables communication with users from all Member States in a standardised workflow, enabling communication in the user's local language. There are 26 nationally localised versions of the tool, including 22 Member States, 2 EFTA and 2 candidate countries. Centralised storage of information (in Comext environment) provides an easy tool to build the knowledge base created from frequently asked questions (FAQ). This is an example of efficiency gain based on cooperation between Eurostat, Member States and their national support centres resulting in high quality user support. Users often request a customised data query which results in a huge data file which is not suitable to be transferred by e-mail. XT-NET/Assist offers a safe environment facilitating transfer (upload/download) of voluminous data files practically without any limitation on size.

The standardised workflow of processing user requests includes creating requests, assignment to the appropriate recipient, follow-up of request status (opened, waiting user input, transferred, pending, resolved, closed) and finally the option to reuse the answer for other similar needs. An important aspect of the tool is improved cooperation between Member States without needing to involve Eurostat.

Future development of the system would enable us to focus on improving the timeliness of requests for validation and correction of data and statistical values. The system will monitor and dispatch all query replies (Eurostat — Member States) related to detailed data quality aspects. The trade statistics domain is the first in Eurostat which fully implemented XT-NET/Assist in practice. In the future the application would also serve as a communication tool between NSIs and Eurostat. Eventually, the same application could be also used to establish a communication network at primary data collection level, i.e. between the providers of statistical information (PSIs) and NSIs.

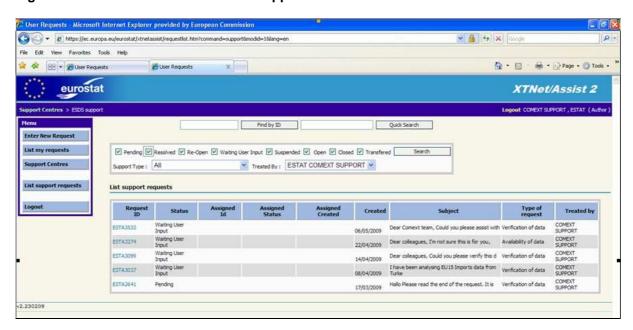


Figure 3: Printscreen of XT-NET/Assist application

2.3 Analysis and Quality

Another step in trade statistics data production includes analytical and quality assessment system. Quality policy of Eurostat is governed by European Statistics Code of Practice [3] with the following leading principles:

- Relevance
- Accuracy and reliability
- Timeliness and punctuality

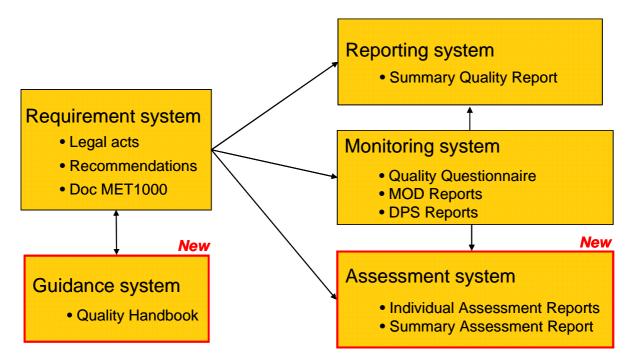
- · Coherence and comparability
- Accessibility and clarity

Quality framework in trade statistics also takes into account additional criteria recommended by the ESS standard for Quality Reports, such as:

Trade-offs between output quality components
Assessment of user needs and perceptions
Performance, cost and respondent burden
Confidentiality, transparency and security

The Quality Framework in European Trade Statistics developed recently in cooperation between Eurostat and Member States is presented in Figure 4 below. Yellow boxes represent 5 main systems which are mutually interlinked. The document labelled as "DOC MET 1000" prepared by Eurostat outlines a set of quality items (indicators, methods and practices) and their associated requirements in terms of target values or minimum standards. The next section gives a closer look at one element of the Monitoring System – the Mirror Outlier Detection (MOD) Report to provide a concrete example of analytical work which can be performed only at the EU level [11].

Figure 4: Quality Framework in European Trade Statistics

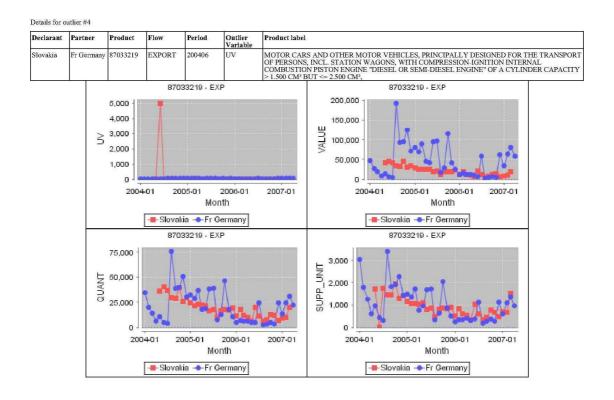


The exports of product X from country A to country B should be in theory equal to imports of country B from country A of the same product. However, since the Intrastat system came into operation in 1993, bilateral comparisons have revealed major and persistent discrepancies in the various Member States' intra-EU trade statistics. The main reasons why "mirror" data flows don't match are different exemption thresholds for Intrastat across MSs, late or non-response, statistical confidentiality applied on one side, misclassifications of products etc. In order to identify these discrepancies, Eurostat systematically develops new tools and methods aiming to improve trade data quality. MOD is a software tool analysing detailed Intrastat data in Comext database. Records of monthly intra-EU trade data, at the CN8 level of product classification can be thought of as time series of records. The fields representing value (in 1000 of euro), quantity (in 100 of kg) or supplementary unit (pieces, cubic metres, etc.) and unit value (EUR/kg) can also be conceived as time series. The analysis is based on detecting outliers in mirror time series of the same product as declared by two different reporting countries (export-dispatch on one side and import-arrival on the other). Each Member State's detected outliers are compared with all other remaining Member States. The method of Median Absolute Deviation (MAD) is used to identify several categories of outliers:

- Green outlier the case when a mirror outlier is found for an outlier data are presumably correct:
- Red or black outlier the mirror time series is conspicuously free of outliers (red) or no corresponding series exists (black) - probably discovered a problem with the quality of the data;
- Pink outlier exists due to statistical confidentiality.

Figure 5 below shows an example of red outlier. In this case an outlier was found in the time series of Slovak exports (red line) of product 87033219 (motor cars) to Germany for the period of January 2004 to February 2007, but no corresponding outlier was found in the mirror series (blue line). Upper right picture indicates total value of exports in euro as declared by Slovakia and at the same time the value of German imports from Slovakia for the same period as declared by Germany. The upper left picture indicates the same series expressed in unit values (UV - euro/pc). In this case the outlier of around 5 millions euro per 1 car was identified in June 2004. Two pictures in the lower part depict corresponding time series of quantities (100 kg) and supplementary units (pieces). This kind of outlier might indicate wrong data entry (price) by Slovak reporter.

Figure 5: Mirror Outlier Detection: Slovak exports (dispatches) to Germany



2. Future developments

The trade statistics domain cannot avoid future challenges of EU statistics as they were recently outlined in Communication of the Commission [2] of a vision on the production method of EU statistics. For many decades, the production of European statistics has been based on a model in which the NSIs of the individual Member States each produce their own national statistics in a particular domain. In order to guarantee comparability and consistency among the data of all Member States, the output from NSIs is harmonised according to agreed standards. The statistics produced by the individual NSIs are compiled by Eurostat so as to obtain European totals.

At present, the business architecture of most of the EU NSIs is still mainly based on a traditional stovepipe model. In such a model, the production of statistics operates through the various production lines of the different statistical domains and every single product stovepipe corresponds to a specific domain of statistics, together with the corresponding production system. Like the trade statistics domain, the whole production process from survey design (Intrastat legislation) over data collection and processing to dissemination (Comext) takes place independently of other domains, and each has its own data suppliers and user groups.

The current stovepipe approach to ESS business architecture is depicted in Figure 6 below. Orange layers of boxes on the left represent MSs with their mutually heterogeneous and internally parallel production systems. Similar situation occurs at Eurostat (blue box) where only quality and metadata management (horizontal box) covers multiple domains.

Regulations incl.

Agreed methodology

To Manage Holes with swappened and state and st

Figure 6: ESS now - stovepipe model.

The overall principle is to move from a stovepipe model to a much more integrated model in the Member States, in Eurostat and also between Eurostat and the Member States. The main disadvantages of stovepipe model are heavy burden on respondents; it is not flexible enough to collect data that cover multiple dimensions (e.g. globalisation); and is inefficient and costly.

All the steps in the statistical production process would be impacted: from identifying user needs in Eurostat, regulations drafted jointly by Eurostat and the Member States, and internal technical processes in the Member States and in Eurostat. Such a possible new model is shown in Figure 7 below. As one can see, the major difference between these two modes lays in different shapes of boxes representing steps of production process (user needs, development and design, collection, etc.). The future model is characterised by integrated multi-dimensional approach to several

production steps design accompanied by the introduction of shared data warehouses and registers placed at various steps of data production either at national or EU level.

Grouping by e.g. statistical unit Same tools, methodologies are used

Figure 7: ESS in the future - integrated model

There are eight main steps to this integration process: Four functional steps are (i) Integrating user needs, (ii) Streamlining regulations, (iii) Converging methodologies, (iv) Harmonising metadata; and four technical steps include (v) Finalising the data life cycle (CVD³) initiative, (vi) Integrating business registers, (vii) Implementing data warehouses and (viii) Harmonising infrastructure and sharing tools.

The concrete example of how to better integrate statistical requirements and to reduce burden on providers of information can be illustrated on business statistics domain. A company with 200 employees produces parts for the automobile industry. At the beginning of each year it submits two surveys used for Structural Business Statistics. They cover turnover, purchases of goods and services, operating surplus, employment, personnel costs and investments. It also submits data on its use of energy for energy statistics. It reports monthly on its trade inside the EU (Intrastat) in value and volume. It also submits monthly reports on its business trends for Short term statistics (turnover, employment, new orders). It reports monthly on its production of goods in value and volume. For each separate data collection, it has to provide the same information on certain basic company features such as turnover.

In an integrated system, many of these data could be obtained from existing administrative data and/or extracted directly from company accounts. For the remainder, one monthly survey should suffice to gather the information that cannot be collected otherwise.

Another example can be taken directly from our trade statistics domain. An integrated approach to user support at ESS level can be achieved by common usage and sharing of application XT-NET/Assist which was described in section 2.2 above.

Finally, it should be noted that Eurostat is committed to assist all ESS members in implementing new vision in the area of business statistics via 5-year grant "Programme for the Modernisation of European Enterprise and Trade Statistics" (MEETS) [12].

³ CVD is Eurostat internal initiative with the objective to streamline IT architecture by using generic and shared components, so called Building Blocks for usage by a wide range of statistical domains.

References

- [1] Berthomieu Anne: Managing of huge administrative data: the example of EU external trade statistics, internal document, Eurostat.
- [2] Communication from The Commission To the European Parliament and the Council on the Production Method of EU Statistics: a Vision for the Next Decade [COM(2009) 404 final]: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:PDF
- [3] Eurostat: *European Statistics Code of Practice:* http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/code_practice.pdf
- [4] Eurostat, 2006: Statistics on the trading of goods User guide. Methods and Nomenclatures, Luxembourg: Office for Official Publications of the European Communities. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-BM-06-001/EN/KS-BM-06-001-EN.PDF
- [5] Eurostat Easy Comext database: http://epp.eurostat.ec.europa.eu/newxtweb
- [6] Eurostat: External and intra-European trade: Statistical yearbook Data 1958-2007: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-GI-08-001/EN/KS-GI-08-001-EN.PDF
- [7] Eurostat: Monthly Bulletin on External and intra-European Union trade: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-AR-09-008/EN/KS-AR-09-008-EN.PDF
- [8] Eurostat: News Release Euro-indicators: http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/6-17082009-AP/EN/6-17082009-AP-EN.PDF
- [9] Eurostat: Panorama of European Union trade -Data 1999-2006: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-DJ-07-001/EN/KS-DJ-07-001-EN.PDF
- [10] Eurostat Pocketbooks: External and intra-European Union trade, Data 2002–07, 2009 edition: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-CV-08-001/EN/KS-CV-08-001-EN.PDF
- [11] Plánovský Ján, Micheal Ashbrook: *The Role of Eurostat in Processing of International Trade Statistics*, Forum Statisticum Slovacum, 5/2008, Bratislava.
- [12] Decision No 1297/2008/EC of the European Parliament and of the Council of 16 December 2008 on a Programme for the Modernisation of European Enterprise and Trade Statistics (MEETS). http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:340:0076:0082:EN:PDF