

C. Data processing and database management at the responsible agency

C. Data processing and database management at the responsible agency

1. *Characteristics of data processing, data flow and data transformations*

8.18. *Characteristics of data processing at the responsible agency in contrast with other statistical activities:* The statistical processing of merchandise trade data involves dealing with large numbers of data sets of relatively simple structure. These data sets are in general obtained from customs declarations and received from customs. Further characteristics are the use of extensive and usually automated validation and quality checking procedures, the storing of processed data and metadata in well-maintained databases capable of performing customized data queries, and timely provision to users of large data sets in various formats. All these activities imply the intensive use of information technology which frequently requires that significant IT resources be specifically dedicated to trade statistics. Particular challenges for statistical data processing can arise when revisions or corrections need to be coordinated and agreed between customs and the responsible agency. A further potential difficulty is the integration of data from other sources, as those data, for example, might not follow the required standard format.

8.19. *Data transformations.* The following data transformations are often executed at the responsible agency: suppression or removal of certain information (due to issues of confidentiality or quality), correction of existing data and supplementation of existing data through estimation or other means (i.e., if certain characteristics are not provided).

Box VIII.1 The statistical production process: example from Italy

The production process carried out by ISTAT encompasses a number of tasks/production stages which ranges from the upload of raw customs data to the release of official data on external trade statistics. In particular, they include:

- (a) Automatic upload or manual data entry of customs data;
- (b) Exclusion of trade flows not relevant for the compilation of external trade statistics;
- (c) Standardization of customs data according to statistical standards, including both classification and analytical variables;
- (d) Rapid detection and revision of major outliers, having significant impact on the aggregate trade figures published as flash estimates or preliminary data;
- (e) Thorough analysis and revision of outliers at the product/country level, including mis-classification problems at the product, country or other statistical variables level;
- (f) Estimation of a possible random item or unit (non-response problems);
- (g) Estimation procedures related to "structural biases" in customs data, such as systematic delays in data transmission, under coverage due to the adoption of exemption thresholds, etc.;
- (h) Estimation of peculiar external trade flows not covered, or poorly covered, by custom data.

These activities require relevant efforts in terms of both hardware and software. The hardware component includes the presence of relevant data storage capability and an appropriate stock of human resources devoted to each stage of the production process. The software component is more intangible but nevertheless crucial for the production of high-quality trade figures. It refers to the stock of knowledge and technical capabilities on data management, data classification and data analysis which is only partially codified in standard IT and statistical procedures; rather, it is mainly embodied in the human capital devoted to the production of trade statistics. It concerns, for example, knowledge on the best way to check, revise and classify specific trade flows, based on an extensive knowledge of product characteristics and feedback from trade operators and external experts.

As a result, the successful management of trade statistics by ISTAT requires the setting up of an appropriate division of labour which takes into account not only the hardware but also the software components of the external trade statistics production process. In particular, it is recommended that:

- (a) An efficient IT framework for the upload and management of customs data be designed, with a dedicated pool of IT technicians;
- (b) A sound methodological approach to outliers detection be designed and implemented;
- (c) A limited pool of specialized clerks with an extensive knowledge of product characteristics needed to manage data-quality problems be established and maintained. In particular, the criteria adopted for assigning a given group of products to each expert should be consistent with human resources constraints and in line with national trade characteristics;
- (d) A risk management approach be adopted to clearly identify critical bottlenecks in the production process which may represent a relevant threat in terms of data quality or timeliness in published trade figures.
- (e) A risk management approach be adopted to clearly identify critical bottlenecks in the production process which may represent a relevant threat in terms of data quality or timeliness in published trade figures.

8.20. *The role of customs.* Custom declarations are the main and usually preferred data source for merchandise trade statistics. Not only are customs authorities providing this information to the responsible agency, but, they have a very strong influence on the quality of the information provided (see chapter IX for details, in particular para. 9.5 on data processing and validation). In this context, it is critical that customs work with the traders or brokers who enter the information to ensure that the data required for statistical purposes are adequately captured in the customs declarations. At the same time, the responsible agency needs to make customs aware of these requirements (see chap.V for details).

2. Examples of data processing systems at the responsible agency

8.21. *Eurotrace software: data- processing software for external trade statistics.* The Eurotrace software, distributed free of charge by Eurostat and implemented in many developing countries,^[15] allows (a) the importation and management of the data necessary to the development of the external trade statistics (in particular the customs data), (b) the treatment of these data, in particular through carrying out quality controls and the application of standards, (c) the working out and calculation of a certain number of aggregates, in particular indices of foreign trade and (d) their export for dissemination and publication. Eurotrace consists of the following separate applications that work together: Eurotrace DBMS, the Eurotrace Data Editor and the Comext Standalone Data Browser.

8.22. *Eurotrace applied in Trinidad and Tobago.*^[16] The Central Statistical Office (CSO) of Trinidad and Tobago has developed a Eurotrace application which has transformed its trade statistics data dissemination. As a result of the implementation of the Eurotrace Trade Statistics application, the time taken to respond to a wide array of ad hoc data requests from international, regional and local data users has been significantly reduced.^[17] Further improvements depend largely on the implementation at customs of ASYCUDA which would replace the current system of manual data capture based on copies of declaration forms. The proposed future data flow will be greatly simplified and will consist of data reception from ASYCUDA, importation to Eurotrace, validation in Eurotrace, upload of validated data and data extraction/direct data download through the Comext Browser.