

C. Possible approaches and solutions when merging data from different sources

C. Possible approaches and solutions when merging data from different sources

7.11. *General considerations for the use of additional data sources.* Compilers need to be aware of the different sources that might be available to provide required information about certain trade transactions that otherwise would not be available. Further, compilers need to gain a thorough understanding of the contents and limitations and the quality of the additional data sources, and obtain adequate access to these data sources. Appropriate institutional arrangements between the compiling agency and the agency responsible for the additional data source need to be in place (see chap. V for details).

7.12. *Merging microlevel data from different sources.* The following steps might be applicable when merging microlevel data from different sources:

- (a) Transform, to the best possible extent, the data from non-customs sources into a standard format that can be readily handled by the IMTS compilation systems;
- (b) Assess the data from non-customs sources, e.g., by comparing it with data from other sources;
- (c) Apply data editing operations, such as re-scaling or estimation of particular data items;
- (d) Add new records to an existing data set or combine records from different sources, including the elimination or correction of existing records, as needed, to avoid any double counting;
- (e) Validate and finalize the combined data set, including, e.g., imputation/estimation of missing quantities.

7.13. *Merging and reconciling data from different sources on the aggregate level.* The additional data sources might not provide sufficient detail to generate data records on the microlevel or might provide only macrolevel information which could be used to establish certain totals (i.e., for commodities or partners). In this case, so-called dummy records, which would represent only a certain value without full commodity or partner detail, could be generated. However, countries might encounter many different situations and might adopt different practices.

7.14. *Supportive measures.* Country experience indicates that certain measures can be taken to facilitate the merging of data from different sources. Compilers may consider:

- (a) Establishing effective controls in the compiling agency to ensure timely replacement of preliminary data from one source by final data obtained from another source (e.g., partner data on a country-of-consignment basis received from customs may be replaced by data on a country-of-last-known-destination basis (for the same goods) received from other governmental agencies, if the latter are judged to be of better quality);
- (b) Developing estimation and imputation procedures to deal with the missing data fields (e.g., estimates of quantities for the current month can be based on current values and on the unit value of the previous month);
- (c) Conducting an ongoing campaign to sensitize customs officers and employees of other source agencies regarding the importance of trade statistics for various purposes;
- (d) Establishing a system-wide terminology-management strategy to ensure the use by all agencies of a consistent terminology in questionnaires. Further, the same classifications for commodities, partner countries, quantity units and modes of transport should be used;
- (e) Running training programmes for staff involved in data compilation (both those of the compiling agency and those of the source agencies, particularly on statistical standards and requirements, conceptual standards and the use of appropriate software) in order to improve staff skills in compiling and merging data from different sources;
- (f) Conducting regular meetings between staff of compiling agencies and staff of source organizations (including staff of large importing and exporting enterprises) to establish stable and efficient working arrangements and complementing such meetings by periodic follow-up phone calls and visits;
- (g) Establishing, to the extent possible, a direct computer link with data suppliers to facilitate data transmission and allow for better and faster verification of incoming data; and using standard classifications and appropriate correlation tables to identify and link the various sets of data;
- (h) Coordinating the installation of computer hardware and software in the compiling and source agencies to ensure their compatibility;