# C. Integration, consolidation and merging of data

## C. Integration, consolidation and merging of data

13.14.    The compilation of data within the statistical framework for describing the international supply of services requires the use of a wide range of data sources, as some sources are complementary and their combined use can result in the production of more detailed and comprehensive statistics.

13.15.    The integration of different sources and the reuse of surveys for several purposes could also reduce the burden for compilers, in particular in the context of compiling modes of supply data. Furthermore, linking data on a microlevel could allow for broader types of analysis of the resulting data. The potential of the approach of using information from other statistical frameworks therefore deserves further exploration by the compiler. For example, in the context of the compilation of inward FATS, it is possible to integrate information from foreign direct investment (FDI) surveys and domestic enterprise surveys. Other examples include using information from sources on migration, tourism, household expenditures, population or taxes (see chapter 15). That information could be used in the compilation of resident/non-resident trade in services data (see chapter 14) or to compile quantitative indicators on modes 2 or 4 (see chapter 16).

13.16.    The consolidation and merging of data sources provides multiple advantages for compilers. First, it improves coverage and the diversity of information.  Second, it usually reduces the resources required to collect the statistics.  Third, it results in statistics of higher quality. However, compilers must be aware of the risk that some of the data sets resulting from the integration of various data sources may be internally inconsistent.

13.17.    Consolidation is a straightforward summation of data from multiple sources in which the sources provide information on non-overlapping parts of the total. Meanwhile, the merging of data sources is usually not a simple task, as compilers must find common denominators between data from multiple sources that cover the same entity or activity. The business register is of vital importance in such linking of microdata.

13.18.    Compilers should take into account all relevant dimensions of data sources before merging; for example, the content of each source could be more or less compatible with a certain services category definition. For such reasons, compilers should have in-depth knowledge of the methodologies and definitions used by the other data sources to ensure consistency and the comparability of the resulting merged data. Various dimensions of the different sources must be analysed by the compilers during merging, most notably: entities covered, services categories or activities identified, the availability of geographical breakdown and the period of reference.

13.19.    **Entities covered**  Sources could collect statistics at different structural levels of the entity.  Some sources may target the legal entity, while others could survey entities responsible for the production. Companies could merge, be acquired, disappear or simply modify their organizational structure so that entities might end up being different from source to source if they have been surveyed at different points in their structural composition. There could also be differences between how the business register defines the structural organization of an entity and how that entity defines its own structural organization. That might result in the statistical compiler unwittingly incorrectly comparing entities from different sources.

13.20.    It is good practice for compilers to analyse the coverage, possible overlap and potentially differing definitions of variables across the different sources available, and determine, on a case-by-case basis,  which data sources are most appropriate to use.

13.21.    **Services categories** Compilers should compare the services categories available from each source to be merged. Details and aggregations provided could be different from source to source. In the absence of clear definitions and guidance, respondents may interpret survey questions incorrectly and include invalid transactions or exclude pertinent ones.

13.22.    **Geographical breakdown**  Data sources could have different levels of geographical breakdown, or even none at all. For example, the country of transaction identified in the ITRS could be different from the country that actually purchased or sold the service.

13.23.    **Period of reference**  Values reported on a monthly basis from one source could differ from the annual value reported by another source (e.g., where there is an annual reconciliation or balancing process). Moreover, difficulties may arise when integrating data from administrative sources on a fiscal year basis with information collected on a calendar year basis.  Similarly, there may be challenges in bringing together data reported on a monthly or quarterly basis with data reported on a through-the-year basis.

13.24.    MSITS 2010 recommends the use of the accrual basis for determining the *time of recording* of transactions. The accrual basis provides the most comprehensive information because all flows are recorded, including non-monetary transactions, imputed transactions and other flows. The change of economic ownership is central in determining the time of recording on an accrual basis.[1]

13.25.    Other factors, such as the threshold for certain sources or the frequency of the survey, may also affect the comparison between sources. If a source is available only once every two years, compilers should develop a strategy to preserve the comparability for the year that the source is not available. All such possible differences between sources may complicate the comparability and integration of data sources,  at both the global level and the microlevel.

13.26.    **Possible approaches and solutions**  There is no quick fix for problems with merging data. Compilers must have in-depth knowledge of the data with which they work and need to recognize the strengths and weaknesses of each source. Data sources are rarely equivalent; one source may provide reliable information for some transactions but be weak for others.  Ideally, if compilers can identify a source in which they have more confidence, that source could serve as a benchmark against which to compare the other sources.

13.27.    **Linking various data sources**  It is vital that compilers avoid approaching businesses multiple times with different surveys covering the same or similar information. Above all, statistical surveys should not request information that the business has already supplied in another data gathering mechanism. The central SBR should be linked to the trade register to enable the analysis of the effects of the international supply of services on production, employment and enterprise performance.[2] The enterprise is the suggested statistical unit to be linked between trade and business statistics; thus, data collected and registered at the level of the declaring unit of trade operators can be aggregated to the level of the whole enterprise via characteristics available in the SBR. The linking of trade and business statistics enables the generation of relevant information on the structure of the international supply of services without collecting additional data from businesses.

[1] See *BPM6 Compilation Guide*, para. 1.21.
[2] See also chapter 16.