B. Metadata: basic concepts and definitions and the role of the Statistical Data and Metadata Exchange

18.11. **Definition of metadata** Metadata are data that define and describe other data and processes.[1] Data become metadata when they are used to describe other data.[2]

18.12. **Scope of statistical metadata** Statistical metadata, according to the Statistical Commission of the United Nations, describe various elements of the statistical processes, including collection, processing and production of statistical data, and indicate the data sources and tools that are instrumental in statistical production, such as statistical standards and classifications, business registers and frames, statistical methods, procedures and software. Section C provides an indicative list of structural and reference metadata items relevant in the context of statistics on the international supply of services.

18.13. **Institutional arrangements for metadata compilation** To reduce the burden associated with projects on metadata for statistics compiled within the framework for describing the international supply of services, it is good practice for compilers to cooperate closely with the specific units within the national statistical system responsible for ensuring that metadata is produced, and for the metadata to adhere to a standard format and be properly maintained and updated.

18.14. The Statistical Commission recommends the use of standard terminology for metadata across the various statistical domains to facilitate the international comparison of data.[3] The Commission is also increasingly encouraging countries to treat metadata compilation and dissemination as integral parts of the statistical process in any statistical domain, and promotes the standardization of the compilation and dissemination of metadata.[4]

18.15. **The way forward: metadata warehousing** Statistical agencies have traditionally developed separate databases for each statistical output. While that practice may simplify development processes, it can hinder the successful integration of statistics, especially if there is no effort to standardize variable definitions, labels and formats. Use of a centralized data warehousing system for data and metadata can make creating, maintaining and accessing metadata more efficient and can contribute to the integration of economic statistics. The process is being facilitated as better information and communications technology tools become available.[5]

18.16. With well-designed data warehouses, the dissemination of data and metadata becomes integrated with the collection and processing components of the statistical production process. A data warehouse should establish a simple and efficient process for accessing data to provide the following:

(a) Comprehensive metadata to facilitate understanding and analysis;
(b) Consistent and coherent long-term time series;
(c) Reliable information about the availability of data;
(d) Information about the availability of updated versions of published series;
(e) Contact details for the people who can provide more information about a statistical output.

18.17. The implementation of a more comprehensive metadata system is an important prerequisite for developing an integrated questionnaire in the statistical system. The metadata will eventually provide the necessary coherence among the various estimates and data collection tools involved in the production of statistical information. For sophisticated users, metadata are not only relevant for concepts related to units, variables and classifications, they are also relevant for the quality of data.

**Role of the Statistical Data and Metadata Exchange**

18.18. The SDMX project was developed by an international consortium[6] for use in data and metadata management. The SDMX information model is applicable for much of the information stored and processed within statistical organizations and its use by such organizations is promoted by the Guidelines on Integrated Economic Statistics of the United Nations.[7]

18.19. The use of the standardized information management model is very important for compilers of statistics on the international supply of services, as various agencies participate in data collection and compilation at different stages of the statistical production process, and the establishment of a standardized data sharing among them results in additional efficiency.

18.20. The development of global DSD, which define the structure for the exchange of data (see section C), by the SDMX consortium and international organizations, enables the broader adoption of the SDMX standard for data collection, exchange and dissemination.
Box 18.1

Statistical Data and Metadata Exchange

1. The Statistical Data and Metadata Exchange (SDMX) is an international cooperative initiative aimed at developing standards and employing more efficient processes for the exchange and sharing of statistical data and metadata among international organizations and their member countries.

2. The rationale of SDMX is the standardization of statistical data and metadata access and exchange. With the ever increasing ease of use of the Internet, the electronic exchange and sharing of data are becoming easier, more frequent and important. This heightens the need for the development of a set of common standards for the exchange and sharing of statistical data and metadata, and for making processes more efficient. As statistical data exchange takes place continuously, the gains to be realized from adopting common standards are considerable, for both data providers and users.

3. The objective is to establish a set of commonly recognized standards, adhered to by all players, making it possible not only to have easy access to statistical data, wherever those data may be, but also to metadata that make the statistics more meaningful and usable. The standards are envisaged to help national organizations fulfill their responsibilities towards users and partners, including international organizations, more efficiently. Among other things, they are seen as facilitating the use of Internet-accessible databases that enable the retrieval of data as soon as they are released. Several quality dimensions can also be improved through the use of SDMX standards, such as timeliness, accessibility, interpretability, coherence and cost effectiveness.

[4] The important role in this respect was played by the ECE publication Terminology on Statistical Metadata, Conference of European Statisticians Statistical Standards and Studies, No. 53. (United Nations publication, Sales No. E.00.II.E.21).
[5] See chapter 21 for more information on the use of information and communications technology (ICT) in the statistical process.
[7] Ibid.