From micro to macro data: ModernStats models for the conceptual modelling of statistical metadata in an interoperability perspective

Speaker: Emanuela Recchini, Italian National Institute of Statistics (Istat)

Author(s):

Abstract

Among the actions undertaken by National Statistical Offices (NSOs) to modernise the production of official statistics and keep pace with an ever-changing world is undoubtedly the move from a traditional area frame-based statistical system to a register-based one. The need to analyse and understand through data phenomena (economic, social, environmental) not only individually but also in their interconnections is pushing NSOs towards overcoming a stovepipe statistical production model in favour of an integrated one. The production of broader, cross-sectional and multidimensional statistical information cannot diregard centralisation and statistical integration of data derived from administrative sources, statistical surveys and new data sources.

One of the challenges that NSOs are facing is the simultaneous processing of data (and thus metadata) from multiple statistical registers with the aim to represent and disseminate the complexity of the multisources output and ensure users maximum flexibility and reuse of data and metadata for further processing with a view to interoperability.

In recent years the National Institute of Statistics (Istat) has developed a suitable methodology which involves the conceptual modelling of metadata of macrodata starting from the decomposition of metadata of registers (micro level). Since it refers to multiple time domains and requires the integration of concepts from different thematic areas, metadata management is particularly complex.

Metadata modelling for the description of data from micro to macro level follows ModernStats models. The idea is to document each aggregate (obtained by statistical synthesis operations directly from the microdata dataset) contributing to the indicator calculation with information on the reference production process (referential metadata) and its structure (structural metadata) so as to build up macrolevel metadata.

Such a methodology ensures traceability of data throughout its entire life cycle, reuse of data and metadata, while improving the quality of the statistical information produced.